

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:23:14 ; Search time 121.9 Seconds  
(without alignments)  
9.379 Million cell updates/sec

Title: US-09-171-432A-41

Perfect score: 103

Sequence: 1 QRLKYAQEELSNEVLPPPRK 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0401.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	20	18 W42924	Immunogenic Hepati
2	103	100.0	25	18 W42930	Immunogenic Hepati
3	103	100.0	366	6 P50230	Sequence of hepati
4	103	100.0	854	6 P50287	Sequence encoded b
5	103	100.0	993	6 P50116	Sequence of Hepati
6	103	100.0	993	6 P50231	Sequence encoded b
7	103	100.0	1077	20 W95559	A partial hepatiti
8	103	100.0	1091	14 R32426	Translated from 5'
9	103	100.0	2227	7 P60066	Sequence of viral
10	103	100.0	2227	11 R05697	Attenuated hepatit
11	103	100.0	2227	18 W34074	Hepatitis A virus

12	103	100.0	2227	21 B18607	Amino acid sequenc
13	103	100.0	2227	21 B18609	Amino acid sequenc
14	98	95.1	2227	21 B18608	Amino acid sequenc
15	94	91.3	839	12 R15629	Capsid region of c
16	48	46.6	444	18 W20624	H. pylori cytoplas
17	45	43.7	20	18 W42923	Immunogenic Hepati
18	44	42.7	914	16 R77274	ORC1 subunit of ye
19	44	42.7	914	18 W22224	S. cerevisiae orig
20	43.5	42.2	1091	10 P98500	Partial sequence e
21	43.5	42.2	3685	10 P90290	Human Duchenne mus
22	43.5	42.2	3685	10 P90373	Sequence encoded b
23	43	41.7	302	21 G06817	Arabidopsis thalia
24	43	41.7	362	21 G06816	Arabidopsis thalia
25	43	41.7	372	21 G06815	Arabidopsis thalia
26	43	41.7	390	20 Y05302	S. aureus protein
27	43	41.7	585	18 W01671	Influenza B/Panama
28	43	41.7	585	20 W75443	Influenza virus B/
29	43	41.7	586	18 W01675	Influenza B/Harbin
30	43	41.7	586	20 W75447	Influenza virus B/
31	43	41.7	589	18 W01672	Influenza B/Nether
32	43	41.7	589	20 W75444	Influenza virus B/
33	43	41.7	592	18 W01674	Influenza A/Shanha
34	43	41.7	592	20 W75446	Influenza virus B/
35	43	41.7	1345	21 B18284	Plasmodium falcipa
36	42.5	41.3	86	21 G03559	Human secreted pro
37	42.5	41.3	530	21 B42529	Human ORFX ORF2293
38	42	40.8	135	21 B43022	Human ORFX ORF2786
39	42	40.8	350	22 B50667	C. elegans UNC-5 p
40	42	40.8	351	22 B50666	C. elegans UNC-5 p
41	42	40.8	445	18 W25768	Human MLN 64. Hom
42	42	40.8	1649	22 B65663	Novel protein kina
43	41.5	40.3	284	21 G14909	Arabidopsis thalia
44	41.5	40.3	285	21 G45427	Arabidopsis thalia
45	41.5	40.3	302	21 G14908	Arabidopsis thalia

ALIGNMENTS

RESULT	1
W42924	
ID	W42924 standard; peptide; 20 AA.
XX	
AC	W42924;
DT	28-APR-1998 (first entry)
XX	
DE	Immunogenic Hepatitis A virus peptide YK-1317.
XX	
KW	Immunogenic peptide; immunogenic epitope; P2A protein;
KW	immune response; antibody.
XX	
OS	Synthetic.
OS	Hepatitis A virus.
PN	W03740147-Al.
XX	
PD	30-OCT-1997.
XX	
PF	18-APR-1997; 97WO-US06891.
XX	
PR	19-APR-1996; 96US-0015644.
XX	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Fields HA, Khudyakov YE;
XX	
DR	WPI; 1997-535831/49.
XX	
PT	Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an
PT	immune response to HAV in a mammal or to detect the presence of
PT	antibodies against HAV in a mammal
XX	



KW Hepatitis A virus assay; antigen; antibody.

OS Hepatitis A virus.

PN WO8501517-A.

XX 11-APR-1985.

PD 27-SEP-1984; 84WO-US01552.

XX 30-SEP-1983; 83US-0537911.

XX (MASI ) MASSACHUSETTS INST TECH.

XX Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;

PI Racanelli VR;

XX WPI: 1985-098846/16.

DR N-PSDB; N50330.

XX New hepatitis A virus cDNA - useful in assays for the virus and  
PT for prodn. of the viral antigen and antibodies to it

XX Example; Fig 7; 60pp; English.

XX The inventors claim HAV cDNA and a method for producing it, whereby  
CC large amts. can be obtd. economically. The cDNA is useful in the  
CC assay for detection of HAV quickly and easily and with high  
CC sensitivity and specificity. The HAV cDNA is also used in the prodn.  
CC of HAV antigen or antibodies to it. The antibodies may be monoclonal.

SQ Sequence 854 AA;

Query Match 100.0%; Score 103; DB 6; Length 854;

Best Local Similarity 100.0%; Pred. No. 1.1e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

Db 810 qrlkyaqeelsnevlppprk 829

RESULT 5

ID P50116

XX P50116 standard; Protein; 993 AA.

AC P50116;

DT 30-SEP-1991 (first entry)

XX Sequence of Hepatitis A virus (HAV) immunogenic peptides  
DE VP-1, VP-2, VP-3 and VP-4.

XX Antigenic protein; immunogen; vaccine.

XX Hepatitis A virus (strain CR326).

XX EP154587-A.

XX 11-SEP-1985.

XX 27-FEB-1985; 85EP-0400369.

XX 02-MAR-1984; 84US-0585818.

XX (MERI ) MERCK & CO INC.

XX Linemeyer DL, Menke JG, Reuben RG, Mitra SW;

XX WPI: 1985-224964/37.

DR N-PSDB; N50139.

XX

PT New nucleotide sequences coding for hepatitis A virus antigens -  
PT useful for eliciting normal immune response and in vaccines for  
PT protecting against the virus

XX Example; Page 11-17; 32pp; English.

XX Within the sequence in N50139 is encoded the information necessary  
CC to make the antigenic proteins of HAV. The sequences encoding for  
CC the structural proteins begin at base 403. The key sub-unit  
CC sequences within vp-1, designated Sequences I, II, III, IV, and V,  
CC start, respectively at 1882, 1963, 1999, 2146, 2347. Other  
CC nucleotide sequences which are valuable as encoding antigenic  
CC proteins are the sequences from base 1749 to base 2722; from base  
CC 1487 to base 2980 and from base 1644 to base 2722. The sequence from  
CC base 1749 to base 2722 is esp. valuable as a vector for producing  
CC antigen protein. Sequences II-V are claimed. X in P50116 denotes the  
CC translation of a stop codon.

XX Sequence 993 AA;

Query Match 100.0%; Score 103; DB 6; Length 993;

Best Local Similarity 100.0%; Pred. No. 1.3e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

Db 946 qrlkyaqeelsnevlppprk 965

RESULT 6

ID P50231

XX P50231 standard; Protein; 993 AA.

AC P50231;

XX 28-NOV-1991 (first entry)

XX Sequence encoded by partial sequence of hepatitis A virus (HAV),  
DE including surface protein (VP-1).

XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
KW diagnostic assay.

XX Hepatitis A virus.

XX Key Location/Qualifiers

FT Protein 628..993

FT /note= "claimed; X denotes translated stop codons  
and unspecified triplets"

XX EP138704-A.

XX 24-APR-1985.

XX 09-OCT-1984; 84EP-0402025.

XX 02-MAR-1984; 84US-0585942.

XX 14-OCT-1983; 83US-0541836.

XX (MERI ) MERCK & CO INC.

XX Hughes JV, Scolnick EM, Tomassini JE;

XX WPI: 1985-100818/17.

DR N-PSDB; N50274.

XX New hepatitis A virus surface protein - useful for binding to  
PT neutralising antibodies to the virus

XX Disclosure; Page 17-23; 49pp; English.

XX VP1 is isolated by solubilisation of the intact virus in an aq.

CC anionic surfactant and a reducing agent. The viral proteins are sepd.  
 CC and the protein of molecular wt. 33000 daltons is sepd.

SQ Sequence 993 AA;

Query Match 100.0%; Score 103; DB 6; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20  
 (|||||)|||||

Db 946 qrlkyaqeelsnevlppprk 965

RESULT 7

ID W95559 standard; Protein; 1077 AA.

AC W95559;

DT 28-APR-1999 (first entry)

DE A partial hepatitis A virus (HAV) protein.

KW Hepatitis A virus protein; HAV; P2 region;

cell-culture-adapted HAV strain; infection; accelerated growth.

OS Hepatitis A virus.

PN US5849562-A.

PD 15-DEC-1998.

PF 06-JUN-1995; 95US-0468926.

XX 06-NOV-1991; 91US-0788262.

PR 30-SEP-1983; 83US-0537911.

PR 27-SEP-1984; 84US-0654942.

PR 06-OCT-1988; 88US-0256135.

PR 06-JUN-1995; 95US-0468926.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI Emerson SU, Purcell RH;

XX WPI; 1999-094412/08.

DR N-PSDB; X01006.

XX Chimeric hepatitis A virus strains - with p2 region from

cell-culture-adapted strain in wild-type genome

XX Disclosure; Fig 7A-L; 36pp; English.

XX The present sequence represents a partial hepatitis A virus (HAV)  
 CC protein. The specification describes a DNA construct consisting  
 CC of a wild-type HAV genome in which the P2 region is replaced by the  
 CC P2 region from a cell-culture-adapted HAV strain. The construct is  
 CC used to demonstrate that mutations in the P2 region of a  
 CC cell-culture-adapted HAV strain are sufficient for establishment of  
 CC infection and accelerated growth in cell culture.

XX Sequence 1077 AA;

Query Match 100.0%; Score 103; DB 20; Length 1077;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20

(|||||)|||||

Db 1033 qrlkyaqeelsnevlppprk 1052

RESULT 8

R32426  
 ID R32426 standard; Protein; 1091 AA.

XX R32426;

DT 10-JUN-1993 (first entry)

XX Translated from 5' region of Hepatitis A Virus genomic clone.

DE HAV HM-175; chronic liver disease; picornavirus.

OS Hepatitis A Virus.

XX Key Location/Qualifiers

XX Region 238..1091

FT /label= ORF

FT /note= "second putative initiation codon at

FT position 240"

FT Region 1..711

FT /note= "X's correspond to nonsense codons,

FT i.e. this region is not an ORF"

XX US7788262-A.

PD 15-DEC-1992.

XX 30-SEP-1983; 83US-0536911.

XX 27-SEP-1984; 84US-0654942.

PR 06-OCT-1988; 88US-0256135.

PR 30-SEP-1983; 83US-0536911.

PR 06-NOV-1991; 91US-0788262.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.

XX Baltimore D, Feinstein SM;

PI Purcell RH, Racaniello VR, Ticehurst JR;

XX WPI; 1993-067429/08.

DR N-PSDB; Q36934.

XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.

XX of antigen and antibodies

XX Disclosure; Fig 7; 65pp; English.

XX HAV virion RNA was extracted from the livers of marmosets which had  
 CC been inoculated with HAV (the HAV had previously been passaged twice  
 CC in marmosets). The RNA was used to prepare ds cDNA clones by  
 CC standard methods. Clones contg. inserts which hybridised to RNA from  
 CC HAV-infected African Green Monkey Kidney cells were selected for  
 CC further analysis. A 7.4kb restriction map (about 99% of the HAV  
 CC genome) was constructed from 5 overlapping inserts. The sequence of  
 CC the first 3 kb (approx.) from the 5'-terminus was determined. An  
 CC amino acid sequence was decoded from the entire clone and an open  
 CC reading frame was identified starting at position 238. A comparison  
 CC of the predicted HAV amino acid sequences with the known capsid  
 CC protein sequences of other picornaviruses (poliovirus, foot and  
 CC mouth disease virus and encephalomyelitis virus) revealed areas of  
 CC local homology.

XX Sequence 1091 AA;

Query Match 100.0%; Score 103; DB 14; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20

(|||||)|||||

Db 1047 qrlkyaqeelsnevlppprk 1066



RESULT 9  
P60066  
ID P60066 standard; Protein; 2227 AA.  
XX AC P60066;  
XX DT 26-JUN-1991 (first entry)  
XX DE Sequence of viral L434 polypeptide encoded by the complete  
DE nucleotide sequence of the HAV genome.  
XX KW Diagnosis; vaccine; passive immunotherapy.  
XX OS Hepatitis A virus.  
XX FH Key Location/Qualifiers  
FT Region 1..245  
FT /label= P1.1A  
FT Region 246..491  
FT /label= 1B  
FT Region 492..836  
FT /label= 1C  
FT Region 837..980  
FT /label= P2.2A  
FT Region 981..1076  
FT /label= 2B  
FT Region 1077..1422  
FT /label= 2C  
FT Region 1423..1484  
FT /label= P3.3A  
FT Region 1485..1507  
FT /label= 3B  
FT Region 1508..1678  
FT /label= 3C  
FT Region 1679..2227  
FT /label= 3D  
XX EP199480-A.  
XX PN  
XX PD 29-OCT-1986.  
XX PF 03-APR-1986; 86EP-0302465.  
XX PR 03-APR-1985; 85US-0719329.  
XX PA (CHIR-) CHIRON CORP.  
XX PI Dina D, Potter SJ, Vannest GA, Caput D;  
XX DR WPI: 1986-286213/44.  
XX DR N-PSDB; N60080.  
XX PT Hepatitis A virus nucleotide sequence and polypeptide - and use  
PT in prodn. of vaccines and diagnostic probes  
XX PS Claim 5; Fig 1; 18pp; English.  
XX CC N60080 and oligonucleotide fragments are useful in detection of  
CC hepatitis A virus; transformed hosts may be used for expression of  
CC polypeptides and fragments useful in vaccines without risk of  
CC infection by the virus or in prodn. of particles which are capable  
CC of inducing immunocompetent B cells for passive immunotherapy. Pref.  
CC epitope is derived from AAs 445-657 or 792-848 of the HAV  
CC polypeptide sequence (P60066).  
XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 103; DB 7; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 810 qrlkyaqeelsnevlppprk 829  
RESULT 10  
R05697  
ID R05697 standard; protein; 2227 AA.  
XX AC R05697;  
XX DT 15-AUG-1990 (first entry)  
XX DE Attenuated hepatitis A virus.  
XX KW Hepatitis A virus; vaccine; attenuated.  
XX OS Hepatitis A virus, strain HM-175.  
XX FH Key Location/Qualifiers  
FT Region 1..23  
FT /label=VP4 = 1A  
FT Region 24..245  
FT /label=VP2 = 1B  
FT Region 246..491  
FT /label=VP3 = 1C  
FT Region 492..791  
FT /label=VP1 = 1D  
FT Region 792..980  
FT /label=2A  
FT Region 981..1087  
FT /label=2B  
FT Region 1088..1422  
FT /label=2C  
FT Region 1423..1496  
FT /label=3A  
FT Region 1497..1519  
FT /label=3B = VPg  
FT Region 1520..1738  
FT /label=3C  
FT Region 1739..2227  
FT /label=3D  
XX US4894228-A.  
XX PN  
XX PD 16-JAN-1990.  
XX PF 12-JUL-1988; 88US-0217824.  
XX PR 12-JUL-1988; 88US-0217824.  
XX PR 12-JUL-1988; 88US-0652967.  
XX PA (USSH ) US DEPT HEALTH & HUMAN.  
XX PI Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstone SM;  
PI Daemer RJ, Gust ID;  
XX WPI: 1990-075557/10.  
XX DR N-PSDB; Q03512.  
XX PT Vaccine against hepatitis A virus infection - comprises novel  
XX attenuated hepatitis A virus strain.  
XX PS Claim 1; Fig 1; 18pp; English.  
XX CC The attenuated HAV is useful for inducing protective immunity against  
CC HAV. This strain (pass 35) differs from the wild type HAV HM-175 by  
CC several nucleotide changes distributed throughout the genome, is  
CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
CC suitable for use as an HAV vaccine. It is noted that not all the changes  
CC are necessary for attenuation and use as a vaccine.  
XX



Query Match 100.0%; Score 103; DB 21; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20  
 |||||  
 Db 810 qrikyaqeelsnevlppprk 829

## RESULT 13

B18609  
 ID B18609 standard; Protein; 2227 AA.

XX AC

XX AC

DT 15-JAN-2001 (first entry)

DE Amino acid sequence of live attenuated Hepatitis A virus 4380.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
 KW HAV 4380.

XX Hepatitis A virus.

OS US6113912-A.

XX 05-SEP-2000.

XX 07-JUN-1995; 95US-0475886.

XX 18-SEP-1992; 92US-0947338.

XX 17-SEP-1993; 93WO-US08610.

XX 10-MAR-1995; 95US-0397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

XX WPI; 2000-586464/55.

XX N-PSDB; A75477.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus  
 PT infection, has modified genome compared to wild type

XX Disclosure; Columns 93-104; 72pp; English.

XX The present sequence is derived from a live attenuated hepatitis A  
 CC virus (HAV) of the invention, designated HAV 4380. The sequence is  
 CC produced by modifying wild type HAV strain HM-174. The HAV of the  
 CC invention are adapted to growth in the human fibroblast-like cell  
 CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain  
 CC appropriate attenuation. It is useful as a live vaccine for prophylaxis  
 CC of hepatitis A in humans and other primates.

XX Sequence 2227 AA;

Query Match 100.0%; Score 103; DB 21; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20  
 |||||  
 Db 810 qrikyaqeelsnevlppprk 829

## RESULT 14

B18608

ID B18608 standard; Protein; 2227 AA.

XX

AC B18608;

XX 15-JAN-2001 (first entry)

XX Amino acid sequence of passage 35 Hepatitis A virus called P-35.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
 KW P-35 virus.

XX Hepatitis A virus.

XX US6113912-A.

XX 05-SEP-2000.

XX 07-JUN-1995; 95US-0475886.

XX 18-SEP-1992; 92US-0947338.

XX 17-SEP-1993; 93WO-US08610.

XX 10-MAR-1995; 95US-0397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

XX WPI; 2000-586464/55.

XX N-PSDB; A75477.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus  
 PT infection, has modified genome compared to wild type

XX Disclosure; Columns 67-78; 72pp; English.

XX The present sequence is derived from passage 35 of a wild type  
 CC hepatitis A virus (HAV) strain HM-174. The resulting virus is  
 CC designated P-35 virus. The sequence is modified to produce HAV which  
 CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
 CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
 CC attenuation. It is useful as a live vaccine for prophylaxis of  
 CC hepatitis A in humans and other primates.

XX Sequence 2227 AA;

Query Match 95.1%; Score 98; DB 21; Length 2227;  
 Best Local Similarity 95.0%; Pred. No. 1.7e-06;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20  
 |||||  
 Db 810 qrikyaqeelsnevlppprk 829

## RESULT 15

R15629

ID R15629 standard; Protein; 839 AA.

XX AC

XX R15629;

XX 17-MAR-1992 (first entry)

XX Capsid region of cyno-HAV isolate CY-145.

XX Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.

XX Cynomolgus monkey hepatitis A virus, isolate CY-145.

XX Key Location/Qualifiers

XX Protein 1..245

XX Protein /label= VP0

XX Protein 246..491

XX Protein /label= VP3

```

FT Protein 492..791
FT /label= VP1
FT 792
FT Protein
FT /label= P2
FT /note= "incomplete"
FT Cleavage-site 245..246
FT Cleavage-site 491..492
FT Cleavage-site 791..792
FT Active-site 315
FT Active-site 593
XX
XX US7678828-A.
XX
XX PD 12-NOV-1991.
XX
XX PF 03-APR-1991; 91US-0678828.
XX
XX PR 03-APR-1991; 91US-0678828.
XX
XX PA (USSH ) US DEPT HEALTH & HUMAN.
XX
XX PI Hainan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;
XX
XX DR WPI; 1991-376737/51.
XX DR N-PSDB; Q15180.
XX
XX PT Hepatitis A virus isolates and DNA - used to prepare vaccines for
XX PT preventing hepatitis A virus infection.
XX
XX PS Disclosure; Fig 3; 23pp; English.
XX
XX CC The sequence was deduced from the nucleotide sequence obt'd. by PCR
XX CC amplification of cyno-HAV viral RNA obt'd. from the stool of a
XX CC cynomolgus monkey with serologically and histologically confirmed
XX CC spontaneous hepatitis A. The sequence differs from the human HAV
XX CC isolate HM175 (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci.
XX CC USA 84, 2497-2501), mainly in the VP3 and VP1 proteins. The Gln-Val
XX CC pair at the VP3-VP1 cleavage site in the human isolate is replaced
XX CC by a Gln-Thr pair in the cyno-HAV. The other two cleavage sites are
XX CC the same. Two residues have been identified as part of the immuno-
XX CC dominant region (see feature table) and are different to those in
XX CC the same position in human HAV. The protein and peptides derived
XX CC from it can be used in the prepn. of vaccines for the prevention of
XX CC HAV infection.
XX CC See also R15056.
XX
XX SQ Sequence 839 AA;

Query Match 91.3%; Score 94; DB 12; Length 839;
Best Local Similarity 85.0%; Pred. No. 2.6e-06;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
Db 809 grfkyareelsneilppprk 828

```

Search completed: April 24, 2001, 16:23:17  
Job time: 135 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 24, 2001, 16:34:59 ; Search time 62.1 Seconds  
(without alignments)  
6.187 Million cell updates/sec.

Title: US-09-171-432A-41  
Perfect score: 103  
Sequence: 1 QRLKYAQEELSNEVLPPPRK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	1091	6	Patent No. 5516630-2
2	103	100.0	2227	4	US-08-475-886-2
3	103	100.0	2227	4	US-08-475-886-6
4	103	100.0	2227	4	US-08-397-232-2
5	103	100.0	2227	4	US-08-397-232-4
6	98	95.1	2227	4	US-08-475-886-4
7	94	91.3	839	1	US-08-087-016-2
8	44	42.7	914	1	US-08-484-105-2
9	44	42.7	914	1	US-08-484-106-2
10	43.5	42.2	113	1	US-08-241-853-15
11	43.5	42.2	113	2	US-08-850-917-15
12	43	41.7	585	2	US-08-453-848-11
13	43	41.7	586	2	US-08-453-848-19
14	43	41.7	589	2	US-08-453-848-13
15	43	41.7	592	2	US-08-453-848-17
16	42	40.8	445	2	US-08-691-814B-6
17	41	39.8	295	2	US-08-454-267-7
18	41	39.8	295	2	US-08-941-319-7
19	41	39.8	295	4	US-09-035-098-7
20	41	39.8	460	2	US-08-933-821-17
21	41	39.8	460	3	US-08-934-494-6
22	41	39.8	460	3	US-08-960-507-17
23	41	39.8	460	3	US-09-143-068-6
24	41	39.8	502	1	US-08-484-840-3
25	41	39.8	502	1	US-08-483-094-3
26	41	39.8	816	2	US-08-267-803B-9
27	41	39.8	1719	2	US-08-459-568-4

28	41	39.8	1719	2	US-08-399-411-4	Sequence 4, Appli
29	41	39.8	1719	3	US-08-516-859A-4	Sequence 4, Appli
30	39	37.9	553	1	US-08-328-322-10	Sequence 10, Appli
31	38	36.9	300	2	US-08-946-528-7	Sequence 7, Appli
32	38	36.9	345	1	US-08-843-993-1	Sequence 1, Appli
33	38	36.9	345	3	US-09-059-520A-1	Sequence 1, Appli
34	38	36.9	345	3	US-09-334-275-1	Sequence 1, Appli
35	38	36.9	580	4	US-09-188-930-307	Sequence 307, App
36	38	36.9	1025	5	PCT-US95-04567-2	Sequence 2, Appli
37	37	35.9	118	2	US-08-417-174-2	Sequence 2, Appli
38	37	35.9	118	2	US-08-231-565A-2	Sequence 2, Appli
39	37	35.9	118	2	US-09-007-961-2	Sequence 2, Appli
40	37	35.9	247	2	US-08-463-911-2	Sequence 2, Appli
41	37	35.9	247	4	US-09-140-804-8	Sequence 8, Appli
42	37	35.9	344	1	US-08-843-993-3	Sequence 3, Appli
43	37	35.9	344	3	US-09-059-520A-3	Sequence 3, Appli
44	37	35.9	344	3	US-09-334-275-3	Sequence 3, Appli
45	37	35.9	374	1	US-08-464-148-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
5516630-2  
; Patent No. 5516630  
; APPLICANT: TICEHURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE,  
; STEPHEN M.; PURCELL, ROBERT H.; RACANELLO, VINCENT R.;  
; BAROUDY, BAHIGE M.  
; TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/788,262  
; FILING DATE: 06-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 256,135  
; FILING DATE: 06-OCT-1988  
; APPLICATION NUMBER: 654,942  
; FILING DATE: 27-SEP-1984  
; APPLICATION NUMBER: 537,911  
; FILING DATE: 30-SEP-1983  
; SEQ ID NO: 2:  
; LENGTH: 1091  
5516630-2

Query Match 100.0%; Score 103; DB 6; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 2.9e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 1047 QRLKYAQEELSNEVLPPPRK 1066

RESULT 2  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2

Query Match      100.0%; Score 103; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match      100.0%; Score 103; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match      100.0%; Score 103; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

Query Match      100.0%; Score 103; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 6
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4
```

Query Match 95.1%; Score 98; DB 4; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 3.9e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPK 20  
|||||:|||||:|||||  
Db 810 ORLKVAQEELSNEVLPPPK 829

## RESULT 7

US-08-087-016-2  
; Sequence 2, Application US/08087016  
; Patent No. 5430135  
; GENERAL INFORMATION:  
; APPLICANT: NAINAN, OMANA V.  
; APPLICANT: MARGOLIS, HAROLD S.  
; APPLICANT: ROBERTSON, BETTY H.  
; APPLICANT: BRINTON, MARGO H.  
; APPLICANT: EBERT, JAMES W.  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,016  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,828  
; FILING DATE: 03-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 839 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 91.3%; Score 94; DB 1; Length 839;  
Best Local Similarity 85.0%; Pred. No. 5.7e-07;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPK 20  
|||:|||||:|||||  
Db 809 ORFKYARELSNEILPPPK 828

## RESULT 8

US-08-484-105-2  
; Sequence 2, Application US/08484105  
; Patent No. 5589341  
; GENERAL INFORMATION:

; APPLICANT: STILLMAN, Bruce  
; APPLICANT: BELL, Stephen P  
; APPLICANT: KOBAYASHI, Ryuji  
; APPLICANT: RINE, Jasper  
; APPLICANT: FOSS, Margit  
; APPLICANT: McNALLY, Francis J  
; APPLICANT: LAURENSEN, Patricia  
; APPLICANT: HERSKOWITZ, Ira  
; APPLICANT: LI, Joachim J  
; APPLICANT: GAVIN, Kimberly  
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,105  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman Ph.D., Richard Aron  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 914 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-484-105-2  
  
Query Match 42.7%; Score 44; DB 1; Length 914;  
Best Local Similarity 30.0%; Pred. No. 46;  
Matches 6; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 ORLKVAQEELSNEVLPPPK 20  
:::|:::|:::|  
Db 310 RKIKLGKDDIDASVQPPPK 329  
  
RESULT 9  
US-08-484-106-2  
; Sequence 2, Application US/08484106  
; Patent No. 5614618  
; GENERAL INFORMATION:  
; APPLICANT: STILLMAN, Bruce  
; APPLICANT: BELL, Stephen P  
; APPLICANT: KOBAYASHI, Ryuji  
; APPLICANT: RINE, Jasper  
; APPLICANT: FOSS, Margit  
; APPLICANT: McNALLY, Francis J  
; APPLICANT: LAURENSEN, Patricia  
; APPLICANT: HERSKOWITZ, Ira  
; APPLICANT: LI, Joachim J  
; APPLICANT: GAVIN, Kimberly  
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,106  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 914 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-106-2

Query Match 42.7%; Score 44; DB 1; Length 914;  
Best Local Similarity 30.0%; Pred. No. 46;  
Matches 6; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Oy 1 ORLKVAQEELSNEVLPPPK 20  
Db 310 RKIKLGKDDIDASVOPPPK 329

RESULT 10  
US-08-853-15  
Sequence 15, Application US/08241853  
Patent No. 5693488  
GENERAL INFORMATION:  
APPLICANT: Fang, Kathy S.  
APPLICANT: Hanafusa, Hidesaburo  
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE  
TITLE OF INVENTION: AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/241,853  
FILING DATE: 12-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-078  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-241-853-15

Query Match 42.2%; Score 43.5; DB 1; Length 113;  
Best Local Similarity 55.6%; Pred. No. 5.7;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

Oy 3 LKVAQEELSNEVLPPPK 20  
Db 7 VKHAQEELPP---PPPK 21

RESULT 11  
US-08-850-917-15  
Sequence 15, Application US/08850917  
Patent No. 5854045  
GENERAL INFORMATION:  
APPLICANT: Fang, Kathy S.  
APPLICANT: Hanafusa, Hidesaburo  
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE  
TITLE OF INVENTION: AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850,917  
FILING DATE: 02-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/241,853  
FILING DATE: 12-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-078  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-850-917-15



Query Match 42.2%; Score 43.5; DB 2; Length 113;  
Best Local Similarity 55.6%; Pred. No. 5.7;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 3 LKYAQEELSNEVLPK 20  
      :||||| :|||  
DB 7 VKHAQEELEPP---PPQK 21

RESULT 12  
US-08-453-848-11  
; Sequence 11, Application US/08453848  
; Patent No. 5858368  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volvovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453.848  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/120.607  
; FILING DATE: 13-SEPT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: MGS101CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404)-873-8794  
; TELEFAX: (404)-873-8795  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 585 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Influenza virus  
; INDIVIDUAL ISOLATE: B/Panama/45/90 rHA  
; FEATURE:  
; NAME/KEY: HA signal peptide  
; LOCATION: 1 to 17  
; FEATURE:  
; NAME/KEY: mature rHA  
; LOCATION: 18 to 568  
US-08-453-848-11

Query Match 41.7%; Score 43; DB 2; Length 585;

Best Local Similarity 60.0%; Pred. No. 41;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORLYAQEELSNEVL 15  
      ||| | :|||  
DB 426 QRLSCAMDELHNEIL 440

RESULT 13  
US-08-453-848-19  
; Sequence 19, Application US/08453848  
; Patent No. 5858368  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volvovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453.848  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/120.607  
; FILING DATE: 13-SEPT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: MGS101CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404)-873-8794  
; TELEFAX: (404)-873-8795  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 586 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Influenza virus  
; INDIVIDUAL ISOLATE: B/Harbin/7/94 rHA  
; FEATURE:  
; NAME/KEY: HA signal peptide  
; LOCATION: 1 to 17  
; FEATURE:  
; NAME/KEY: mature rHA  
; LOCATION: 18 to 569  
US-08-453-848-19

Query Match 41.7%; Score 43; DB 2; Length 586;  
Best Local Similarity 60.0%; Pred. No. 41;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORLKAYAOEELSNEVL 15  
||||| :||| ||:|  
Db 427 ORLSGAMDELHNEIL 441

RESULT 14  
US-08-453-848-13  
; Sequence 13, Application US/08453848  
; Patent No. 5858368  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volvovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,848  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/120,607  
; FILING DATE: 13-SEPT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: MGS101CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404)-873-8794  
; TELEFAX: (404)-873-8795  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 589 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Influenza virus  
; INDIVIDUAL ISOLATE: B/Netherlands/13/94 rHA  
; FEATURE:  
; NAME/KEY: AcNPV 61K protein signal sequence  
; LOCATION: 1 to 18  
; FEATURE:  
; NAME/KEY: mature rHA  
; LOCATION: 19 to 571  
US-08-453-848-13

Query Match 41.7%; Score 43; DB 2; Length 589;  
Best Local Similarity 60.0%; Pred. No. 41;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORLKAYAOEELSNEVL 15  
||||| :||| ||:|  
Db 430 ORLSGAMDELHNEIL 444

RESULT 15  
US-08-453-848-17  
; Sequence 17, Application US/08453848  
; Patent No. 5858368  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volvovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,848  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/120,607  
; FILING DATE: 13-SEPT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: MGS101CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404)-873-8794  
; TELEFAX: (404)-873-8795  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 592 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Influenza virus  
; INDIVIDUAL ISOLATE: B/Shanghai/4/94 rHA  
; FEATURE:  
; NAME/KEY: AcNPV 61K protein signal peptide  
; LOCATION: 1 to 18  
; FEATURE:  
; NAME/KEY: mature rHA  
; LOCATION: 19 to 574  
US-08-453-848-17

Query Match 41.7%; Score 43; DB 2; Length 592;  
Best Local Similarity 60.0%; Pred. No. 41;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QRLKYAOEELSNEVL 15  
| | | | : | | | | |  
Db 433 QRLSGAMDELHNEIL 447

Search completed: April 24, 2001, 16:35:00  
Job time: 531 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:25:46 ; Search time 73.14 Seconds  
(without alignments)  
18.792 Million cell updates/sec

Title: US-09-171-432a-41

Perfect score: 103

Sequence: 1 ORLKVAQEELSNEVLPPPRK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_67.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	341	2 S04137	genome polyprotein
2	103	100.0	852	1 GNNYHA	genome polyprotein
3	103	100.0	1358	2 A03905	genome polyprotein
4	103	100.0	2227	1 GNNYHM	genome polyprotein
5	103	100.0	2227	1 GNNYHR	genome polyprotein
6	103	100.0	2227	1 GNNYHB	genome polyprotein
7	98	95.1	2227	1 GNNYMK	genome polyprotein
8	97	94.2	2230	1 GNNYSA	genome polyprotein
9	94	91.3	839	1 GNNYS2	genome polyprotein
10	56	54.4	346	2 S74448	regulatory protein
11	52	50.5	859	2 T43701	DNA-directed RNA p
12	51	49.5	1119	2 T50995	related to cytoske
13	49.5	48.1	443	2 E82046	proteinase HslVU,
14	49	47.6	55	2 P00433	genome polyprotein
15	49	47.6	56	2 P00434	genome polyprotein
16	49	47.6	56	2 P00428	genome polyprotein
17	49	47.6	56	2 P00427	genome polyprotein
18	49	47.6	56	2 P00429	genome polyprotein
19	49	47.6	56	2 P00432	genome polyprotein
20	49	47.6	56	2 P00430	genome polyprotein
21	49	47.6	1174	2 S28976	DNA-directed RNA p
22	48	46.6	442	2 A71969	probable histidine
23	48	46.6	1176	2 A27826	DNA-directed RNA p
24	48	46.6	1191	2 S65068	DNA-directed RNA p
25	47	45.6	592	1 LLBY	actin-binding prot
26	47	45.6	6642	2 T29757	protein UNC-89 - C
27	46	44.7	338	2 T56893	transcription fact
28	46	44.7	480	1 FWPUB1B	ILS globulin beta
29	46	44.7	483	2 T01053	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S04137

genome polyprotein - human hepatitis A virus (strain LCDC-1) (fragment)

C:Species: human hepatitis A virus

C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000

C:Accession: S04137

R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.

Nucleic Acids Res. 17, 3594, 1989

A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A vir

A:Reference number: S04137; MUID:89263805

A:Accession: S04137

A:Molecule type: mRNA

A:Residues: 1-341 <AND>

A:Cross-references: EMBL:X14666; NID:962301; PIDN:CAA32794.1; PID:94377576

C:Genetics:

A:Gene: VP1

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; polyprotein

F:2-340/Product: coat protein ID (VP1) #status predicted <MAT>

Query Match 100.0%; Score 103; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ORLKVAQEELSNEVLPPPRK 20

Db 314 ORLKVAQEELSNEVLPPPRK 333

##### RESULT 2

GNNYHA

genome polyprotein - human hepatitis A virus (strain CR326) (fragment)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A

C:Species: human hepatitis A virus

A:Note: host Homo sapiens (man)

C>Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999

C:Accession: A03904

R:Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra,

J. Virol. 54, 247-255, 1985

A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.

A:Reference number: A03904; MUID:85185648

A:Accession: A03904

A:Molecule type: genomic RNA

A:Residues: 1-852 <LIN>

A:Cross-references: EMBL:M10033; NID:9329592; PIDN:AAA45470.1; PID:9329593

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; polyprotein

F:1-245/Product: coat protein 1A #status predicted <CIA>

F:246-451/Product: coat protein 1B #status predicted <CIB>

F:492-836/Product: coat protein 1C #status predicted <CIC>

F:837-852/Product: core protein 2A (fragment) #status predicted <C2A>

cycloartenol synth  
DNA-directed RNA p  
hypothetical prote  
genome polyprotein  
hemagglutinin HA2  
phosphoglycerate m  
probable phosphogl  
probable potassium  
hypothetical prote  
ORC1 protein - yea  
hypothetical prote  
dystrophin - mouse  
dystrophin, muscle  
protein YGL006w-a  
genome polyprotein  
alkaline-shock pro

```

Query Match      100.0%; Score 103; DB 1; Length 852;
Best Local Similarity 100.0%; Pred. No. 3.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
|||||
Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 3
A03905
genome polyprotein (version 2) - human hepatitis A virus (fragments)
N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein
C;Species: human hepatitis A virus
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 15-Nov-1996
C;Accession: A03905
F;Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feinstone
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985
A;Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA
A;Reference number: A03905; MUID:85166289
A;Accession: A03905
A;Molecule type: genomic RNA
A;Residues: 1-1358 <BAR>
C;Superfamily: hepatitis A virus genome polyprotein
C;Keywords: coat protein; core protein; polyprotein
F;1-245/Product: coat protein 1A #status predicted <CIA>
F;246-491/Product: coat protein 1B #status predicted <CLB>
F;492-836/Product: coat protein 1C #status predicted <CLC>
F;837-854/Product: core protein 2A (fragment) #status predicted <C2A>
F;855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match      100.0%; Score 103; DB 2; Length 1358;
Best Local Similarity 100.0%; Pred. No. 6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
|||||
Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 4
GNHYH
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)
N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro
B; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D
C;Species: human hepatitis A virus
A;Note: host Homo sapiens (man)
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
C;Accession: A25981
F;Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.
J. Virol. 61, 50-59, 1987
A;Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with di
A;Reference number: A25981; MUID:87061253
A;Accession: A25981
A;Molecule type: genomic RNA
A;Residues: 1-2227 <COH>
A;Cross-references: EMBL:M144707; NID:g329582; PIDN:AAA45465.1; PID:g329583
C;Superfamily: hepatitis A virus genome polyprotein
C;Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans
F;1-23/Product: coat protein 1A #status predicted <VP4>
F;24-245/Product: coat protein 1B #status predicted <VP2>
F;246-491/Product: coat protein 1C #status predicted <VP3>
F;492-791/Product: coat protein 1D #status predicted <VP1>
F;792-980/Product: core protein 2A #status predicted <C2A>
F;981-1087/Product: core protein 2B #status predicted <C2B>
F;1088-1422/Product: core protein 2C #status predicted <C2C>
F;1423-1496/Product: protein 3A #status predicted <C3A>
F;1497-1519/Product: protein 3B #status predicted <C3B>
F;1730-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>
F;1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

```

A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:g222597; PIDN:BAA00766.1; PID:g222598  
R:Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure  
A:Reference number: JQ1080; MUID:91311420  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizh  
submitted to the EMBL data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:g61971; PIDN:CAA33490.1; PID:g930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizh  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 94.2% Score 97; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 9.3e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
||||| |||||||  
Db 814 QRLKYAMEELSNEILPPPRK 833

RESULT 9  
GNYS2  
N:Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat prot  
C:Species: simian hepatitis A virus  
A:Note: host Macaca fascicularis (cynomolgus macaque)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
C:Accession: JQ1180  
R:Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A:Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus  
A:Reference number: JQ1180; MUID:91311421  
A:Accession: JQ1180  
A:Molecule type: genomic RNA  
A:Residues: 1-839 <NA1>  
A:Cross-references: GB:M59286; NID:g329599; PIDN:AAA5473.1; PID:g555083  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; glycoprotein; polyprotein  
F:1-23/Product: coat protein 1A #status predicted <VP0>  
F:24-245/Product: coat protein 1B #status predicted <VP3>  
F:246-491/Product: coat protein 1C #status predicted <VP1>  
F:492-839/Product: core protein 2A (fragment) #status predicted <P2p>  
F:261,312,728,756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 91.3% Score 94; DB 1; Length 839;

Best Local Similarity 85.0%; Pred. No. 9.2e-07;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPK 20  
Db 809 QRFKYARELSNEILLPPPK 828  
|||:|||||:|||||

RESULT 10  
S74448  
regulatory protein pcrR - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: protein sll1408  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 18-Aug-2000  
C:Accession: S74448  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
s.  
A:Reference number: S743322; MUID:97061201  
A:Accession: S74448  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-346 <KAN>  
A:Cross-references: EMBL:D90899; GB:AB001339; NID:gl1651650; PIDN:BAAL16600.1; PID:gl165167  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Gene: pcrR  
C:Superfamily: *Synechocystis* regulatory protein pcrR  
C:Keywords: transcription regulation

Query Match 54.4%; Score 56; DB 2; Length 346;  
Best Local Similarity 61.1%; Pred. No. 0.34;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPP 18  
Db 241 ERVKYAEILVKQVLDP 258  
:|||||:|:|

RESULT 11  
T43701  
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain [imported] - *Caenorhabditis el*  
C:Species: *Caenorhabditis elegans*  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 02-Sep-2000  
C:Accession: T43701  
R:Sidow, A.; Thomas, W.K.  
Curr. Biol. 4, 596-603, 1994  
A:Title: A molecular evolutionary framework for eukaryotic model organisms.  
A:Reference number: 222636; MUID:95041334  
A:Accession: T43701  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-859 <SID>  
A:Cross-references: EMBL:U10333; NID:g520506; PIDN:AAA50224.1; PID:g520507  
C:Superfamily: DNA-directed RNA polymerase 132K polypeptide  
C:Keywords: nucleotidyltransferase

Query Match 50.5%; Score 52; DB 2; Length 859;  
Best Local Similarity 62.5%; Pred. No. 4.1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPP 16  
Db 141 QRIKYAREILQKLELP 156  
|||:|:|:|

RESULT 12  
T50995

related to cytoskeleton assembly control protein sLAL [imported] - *Neurospora crassa*  
N:Alternate names: protein B7F18.140  
C:Species: *Neurospora crassa*  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: T50995  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
submitted to the Protein Sequence Database, July 2000  
A:Reference number: 225286  
A:Accession: T50995  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1119 <SCH>  
A:Cross-references: EMBL:AI389891; GSPDB:GN00116; NCSP:B7F18.140  
A:Experimental source: BAC clone B7F18; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B7F18.140  
A:Map position: 6  
A:Introns: 66/3; 123/2; 495/1

Query Match 49.5%; Score 51; DB 2; Length 1119;  
Best Local Similarity 55.6%; Pred. No. 7.9; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPP 18  
Db 165 QRSYASEDYENVRSP 182  
|||:|:|:|

RESULT 13  
E82046  
proteinase HslVU, ATPase subunit HslU VC2674 [imported] - *Vibrio cholerae* (group O1 s  
C:Species: *Vibrio cholerae*  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 17-Nov-2000  
C:Accession: E82046  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833  
A:Accession: E82046  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-443 <HEI>  
A:Cross-references: GB:AE004333; GB:AE003852; NID:g9657266; PIDN:AAF95815.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2674  
A:Map position: 1  
C:Superfamily: heat shock protein hslU; Ftsh/SEC18/CDC48-type ATP-binding domain homo

Query Match 48.1%; Score 49.5; DB 2; Length 443;  
Best Local Similarity 41.7%; Pred. No. 4.8; Mismatches 6; Indels 5; Gaps 1;

Qy 1 QRLKYAQEELSNE-----VLPPPR 19  
Db 117 EKVRFRAEELAEERVLDP 140  
:::|:|:|:|

RESULT 14  
PQ0433  
genome polyprotein - human hepatitis A virus (strain PA21) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0433  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wi  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different ge  
A:Reference number: PQ0427; MUID:92300330  
A:Accession: PQ0433



Search completed: April 24, 2001, 16:25:51  
Job time: 285 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:33:55 ; Search time 38.74 Seconds  
(without alignments)  
17.685 Million cell updates/sec

Title: US-09-171-432A-41

Perfect score: 103  
Sequence: 1 QRLKYAQEELSNEVLPPPRK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	341	1 POLG_HPAV1	P13672 hepatitis a
2	103	100.0	808	1 POLG_HPAVG	Q02381 hepatitis a
3	103	100.0	852	1 POLG_HPAVC	P06442 hepatitis a
4	103	100.0	2226	1 POLG_HPAV2	P26580 hepatitis a
5	103	100.0	2226	1 POLG_HPAV4	P26581 hepatitis a
6	103	100.0	2226	1 POLG_HPAV8	P26582 hepatitis a
7	103	100.0	2227	1 POLG_HPAVH	P08617 hepatitis a
8	103	100.0	2227	1 POLG_HPAVL	P06441 hepatitis a
9	103	100.0	2227	1 POLG_HPAVM	P13901 hepatitis a
10	97	94.2	2230	1 POLG_HPAVS	P14553 simian hepa
11	94	91.3	839	1 POLG_HPAVT	P31788 simian hepa
12	52	50.5	1193	1 RPB2_CAEEL	Q10578 caenorhabdi
13	49.5	48.1	443	1 HSLU_VIBCH	Q9knq7 vibrio chol
14	49	47.6	1174	1 RPB2_HUMAN	P30876 homo sapien
15	48	46.6	1176	1 RPB2_DROME	P08266 drosophila
16	48	46.6	1191	1 RPB2_LYCES	Q42877 lycopersico
17	47	45.6	592	1 ABP1_YEAST	P15891 saccharomyc
18	46	44.7	338	1 TAP4_HUMAN	Q01664 homo sapien
19	46	44.7	480	1 L1SB_CUCMA	P13744 cucurbita m
20	45	43.7	733	1 HEXA_BIADI	Q17127 blaberus di
21	44	42.7	303	1 PMG3_YEAST	Q12326 saccharomyc
22	44	42.7	510	1 PMGI_ANTSP	Q06464 antithamio
23	44	42.7	914	1 ORC1_YEAST	P54784 saccharomyc
24	43.5	42.2	3678	1 DMD_MOUSE	P11531 mus musculu
25	43.5	42.2	3685	1 DMD_HUMAN	P11532 homo sapien
26	43	41.7	268	1 YHC6_YEAST	P38740 saccharomyc
27	43	41.7	574	1 HEMA_INBMD	P03461 influenza b
28	43	41.7	576	1 HEMA_INBDS	P09766 influenza b
29	43	41.7	578	1 HEMA_INBME	P09765 influenza b
30	43	41.7	578	1 HEMA_INBVI	P09767 influenza b
31	43	41.7	583	1 HEMA_INBEN	P10757 influenza b
32	43	41.7	583	1 HEMA_INBOR	P03464 influenza b
33	43	41.7	583	1 HEMA_INBSI	P03463 influenza b

34 43 41.7 585 1 HEMA\_INBBE P17504 influenza b  
35 43 41.7 585 1 HEMA\_INBVK P22092 influenza b  
36 43 41.7 1146 1 AS10\_YEAST P48361 saccharomyc  
37 43 41.7 1210 1 RPB2\_SCHPO Q02061 schizosacch  
38 42.5 41.3 2278 1 FAB1\_YEAST P34756 saccharomyc  
39 42 40.8 214 1 DEVS\_MYXXA Q07766 myxococcus  
40 42 40.8 276 1 YJIC\_ECOLI P39374 escherichia  
41 42 40.8 445 1 ML64\_HUMAN Q14849 homo sapien  
42 42 40.8 554 1 PEX2\_PODAN P51021 podospira a  
43 42 40.8 1005 1 DPOL\_VARV P33793 variola vir  
44 42 40.8 1006 1 DPOL\_VACCC P20509 vaccinia vi  
45 42 40.8 1006 1 DPOL\_VACCV P06856 vaccinia vi

## ALIGNMENTS

RESULT 1  
POLG\_HPAV1  
ID POLG\_HPAV1 STANDARD; PRT; 341 AA.  
AC P13672;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP3; CORE PROTEIN P2A] (FRAGMENT).  
DE P2A]  
OS Hepatitis A virus (strain LCDC-1).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_Taxid=12093;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89263805; PubMed=2542903;  
RA Andonov A.P.; Lau P.; Chaudhary R.;  
RT "Nucleotide sequence of the VP1 gene from a Chinese strain of  
RT hepatitis A virus (HAV).";  
RL Nucleic Acids Res. 17:3594-3594(1989).  
CC -! SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -! PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; X14666; CAA32794.1; -.  
DR PIR; S04137; S04137.  
KW Polypeptide; Coat protein; Core protein.  
FT NON\_TER 1 1  
FT CHAIN <1 1 COAT PROTEIN VP3 (1C).  
FT CHAIN 2 340 COAT PROTEIN VP1 (1D).  
FT CHAIN 341 >341 CORE PROTEIN P2A.  
FT NON\_TER 341 341  
SQ SEQUENCE 341 AA; 38003 MW; 066918289BF126D5 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 341;  
Best Local Similarity 100.0%; Pred. No. 9e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||||  
Db 314 QRLKYAQEELSNEVLPPPRK 333

RESULT 2  
POLG\_HPAVG  
ID POLG\_HPAVG STANDARD; PRT; 808 AA.

```
AC Q02381;
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEIN
DE P2A] (FRAGMENT).
OS Hepatitis A virus (strain CA76).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=31706;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92260183; PubMed=1316423;
RA Khanna B., Spelbring J.E., Innis B.L., Robertson B.H.;
RT "Characterization of a genetic variant of human hepatitis A virus.";
RL J. Med. Virol. 36:118-124(1992).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M6695; AAA5477.1; -
KW Polyprotein; Coat protein; Core protein.
FT CHAIN <1 2 COAT PROTEIN VP4 (P1A).
FT CHAIN 3 223 COAT PROTEIN VP2 (P1B).
FT CHAIN 224 470 COAT PROTEIN VP3 (P1C).
FT CHAIN 471 770 COAT PROTEIN VP1 (P1D).
FT CHAIN 771 >808 CORE PROTEIN P2A.
FT NON_TER 808
SQ SEQUENCE 808 AA; 90632 MW; D80CE7E57A479C12 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 808;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
DB 789 QRLKYAQEELSNEVLPPPRK 808

RESULT 3
POLG_HPAVC STANDARD; PRT; 852 AA.
AC P06442; Q83741; Q83742;
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEIN
DE P2A] (FRAGMENT).
OS Hepatitis A virus (strain CR326).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85185648; PubMed=2985793;
RA Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,
RA Young A., Mitra S.W.;
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA.";
RL J. Virol. 54:247-255(1985).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
```

```
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M10033; AAA5470.1; -
DR PIR; A03904; GNNYHA.
KW Polyprotein; Coat protein; Core protein.
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).
FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).
FT CHAIN 837 >852 CORE PROTEIN P2A.
FT NON_TER 852
SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;

Query Match 100.0%; Score 103; DB 1; Length 852;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 4
POLG_HPAV2 STANDARD; PRT; 2226 AA.
AC P26580;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 24a).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromeans T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination.";
RL J. Virol. 65:2056-2065(1991).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59810; AAA5468.1; -
DR MEROPS; C03.005; -
DR InterPro; IPR000605; -
DR InterPro; IPR001205; -
```

```

DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 794
FT CHAIN 795 900
FT CHAIN 901 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1495
FT CHAIN 1496 1518
FT CHAIN 1519 1737
FT CHAIN 1738 2226
FT CHAIN 2226 AA; 251152 MW; 6CD85A91D684E2BF CRC64;
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D684E2BF CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 7.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 5
POLG_HP4V4 STANDARD; PRT; 2226 AA.
AC P26581;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)]
OS Hepatitis A virus (strain 43c).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12095;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromean T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M59809; AAA45469.1;
CC MEROPS: C03.005;
CC InterPro: IPR001205;
CC Pfam: PF00680; RNA_dep_RNA_pol; 1.
CC Pfam: PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 794
FT CHAIN 795 900
FT CHAIN 901 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1495
FT CHAIN 1496 1518
FT CHAIN 1519 1737
FT CHAIN 1738 2226
FT CHAIN 2226 AA; 251152 MW; 6CD85A91D684E2BF CRC64;
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D684E2BF CRC64;

```

```

FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 794
FT CHAIN 795 900
FT CHAIN 901 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1495
FT CHAIN 1496 1518
FT CHAIN 1519 1737
FT CHAIN 1738 2226
FT CHAIN 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 7.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 6
POLG_HP4V8 STANDARD; PRT; 2226 AA.
AC P26582;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)]
OS Hepatitis A virus (strain 18f).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromean T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M59808; AAA45467.1;
CC MEROPS: C03.005;
CC InterPro: IPR000605;
CC InterPro: IPR001205;
CC Pfam: PF00680; RNA_dep_RNA_pol; 1.
CC Pfam: PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 794
FT CHAIN 795 900
FT CHAIN 901 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1495
FT CHAIN 1496 1518
FT CHAIN 1519 1737
FT CHAIN 1738 2226
FT CHAIN 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;

```

```
FT CHAIN 1088 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.
SQ SEQUENCE 2226 AA; 251292 MW; 24964A63396C8D6B CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 7.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20
    |||||
Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 7
POLG_HPAAVH STANDARD; PRT; 2227 AA.
AC P08617; P06443; O81082;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain HM-175).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILD TYPE;
RX MEDLINE=87061253; PubMed=3023706;
RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,
RA Baroudy B.M.;
RT "Complete nucleotide sequence of wild-type hepatitis A virus:
RT comparison with different strains of hepatitis A virus and other
RT picornaviruses.";
RL J. Virol. 61:50-59(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATTENUATED;
RX MEDLINE=87175701; PubMed=3031686;
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,
RA Purcell R.H.;
RT "Complete nucleotide sequence of an attenuated hepatitis A virus:
RT comparison with wild-type virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).
RN [3]
RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.
RX MEDLINE=85166289; PubMed=2984684;
RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,
RA Purcell R.H., Feinstone S.M.;
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid
RT proteins and RNA polymerase.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -!- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT
CC SHOWN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC
```

```
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M14114; AAA45475.1; -.
CC EMBL; M14707; AAA45465.1; -.
CC EMBL; M14707; AAA45466.1; ALT_INIT.
CC EMBL; M16632; AAA45471.1; -.
CC PIR; A25981; GNNYHM.
CC PIR; A25914; GNNYMK.
CC PIR; A03905; A03905.
CC MEROPS; C03.005; -.
CC InterPro; IPR001205; -.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC CHAIN 1 23
FT CHAIN 24 245 COAT PROTEIN VP4 (PIA).
FT CHAIN 246 491 COAT PROTEIN VP2 (PIB).
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).
FT CHAIN 837 980 CORE PROTEIN P2A.
FT CHAIN 981 1087 CORE PROTEIN P2B.
FT CHAIN 1088 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1496 PROBABLE PROTEIN P3A.
FT CHAIN 1497 1519 PROBABLE PROTEIN P3B.
FT CHAIN 1520 1738 PROBABLE PROTEIN P3C.
FT CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.
FT VARIANT 77 77 K -> R (IN ATTENUATED STRAIN).
FT VARIANT 764 764 E -> V (IN ATTENUATED STRAIN).
FT VARIANT 821 821 N -> S (IN ATTENUATED STRAIN).
FT VARIANT 1052 1052 A -> V (IN ATTENUATED STRAIN).
FT VARIANT 1062 1062 G -> A (IN ATTENUATED STRAIN).
FT VARIANT 1118 1118 K -> M (IN ATTENUATED STRAIN).
FT VARIANT 1151 1151 E -> K (IN ATTENUATED STRAIN).
FT VARIANT 1163 1163 F -> S (IN ATTENUATED STRAIN).
FT VARIANT 1277 1277 V -> I (IN ATTENUATED STRAIN).
FT VARIANT 1500 1500 H -> Y (IN ATTENUATED STRAIN).
FT VARIANT 1805 1805 D -> N (IN ATTENUATED STRAIN).
FT VARIANT 1930 1930 S -> T (IN ATTENUATED STRAIN).
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20
    |||||
Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 8
POLG_HPAAVH STANDARD; PRT; 2227 AA.
AC P06441;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain LA).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12099;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85190549; PubMed=2986127;
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,
RA Merryweather J., van Nest G., Dina D.;
RT "Primary structure and gene organization of human hepatitis A virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).
```

```

CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M20273; AAA45474.1; -
CC PIR; JS0303; GNNYHB.
CC
CC DR MEROPS; C03.005; -
CC
CC DR InterPro; IPR000605; -
CC
CC DR InterPro; IPR001205; -
CC
CC DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC
CC DR Pfam; PF00910; RNA_helicase; 1.
CC
CC KW Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC
CC FT CHAIN 1 23
CC FT CHAIN 24 245
CC FT CHAIN 246 491
CC FT CHAIN 492 836
CC FT CHAIN 837 980
CC FT CHAIN 981 1076
CC FT CHAIN 1077 1422
CC FT CHAIN 1423 1484
CC FT CHAIN 1485 1507
CC FT CHAIN 1508 1678
CC FT CHAIN 1679 2227
CC SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
DB 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 9
POLG_HPAVM STANDARD; PRT: 2227 AA.
AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;
AC Q81090; Q81091; Q81092; Q81093;
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.48)].
OS Simian hepatitis A virus (strain MBB).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88045071; PubMed=2823500;
RA Paul A.V., Tada H., der Helm K., Wissel T., Kiehn R., Wimmer E.,
RA Deinhardt F.;
RT "The entire nucleotide sequence of the genome of human hepatitis A
RT virus (isolate MBB).";
RL Virus Res. 8:153-171(1987).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC

```

```

CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.
CC
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M20273; AAA45474.1; -
CC PIR; JS0303; GNNYHB.
CC
CC DR MEROPS; C03.005; -
CC
CC DR InterPro; IPR000605; -
CC
CC DR InterPro; IPR001205; -
CC
CC DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC
CC DR Pfam; PF00910; RNA_helicase; 1.
CC
CC KW Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
CC
CC FT CHAIN 1 23
CC FT CHAIN 24 245
CC FT CHAIN 246 491
CC FT CHAIN 492 836
CC FT CHAIN 837 980
CC FT CHAIN 981 1087
CC FT CHAIN 1088 1422
CC FT CHAIN 1423 1496
CC FT CHAIN 1497 1519
CC FT CHAIN 1520 1738
CC FT CHAIN 1739 2227
CC SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
DB 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 10
POLG_HPAVS STANDARD; PRT: 2230 AA.
AC P14553;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.48)].
OS Simian hepatitis A virus (strain AGM-27).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12102;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91311420; PubMed=1649901;
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,
RA Purcell R.H.;
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome
RT structure and growth in cell culture with other HAV strains.";
RL J. Gen. Virol. 72:1677-1683(1991).
RN [2]
RP SEQUENCE OF 1750-2164 FROM N.A.
RX MEDLINE=89232168; PubMed=2541023;
RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;
RT "Variations in genome fragments coding for RNA polymerase in human
RT and simian hepatitis A viruses.";

```

```

RL FEBS Lett. 247:425-428(1989).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00924; BAA00766.1; -.
DR EMBL; X15461; CAA33490.1; -.
DR PIR; A30470; GNNYSA.
DR PIR; S04885; S04885.
DR MEROPS; C03.005; -.
DR InterPro; IPR000605; -.
DR InterPro; IPR001205; -.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
DR Pfam; PF0910; RNA_helicase; 1.
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
KW Polyprotein; Coat protein; Core protein; Transferase.
FT CHAIN 1 27 COAT PROTEIN VP4 (PIA).
FT CHAIN 28 249 COAT PROTEIN VP2 (PIB).
FT CHAIN 250 495 COAT PROTEIN VP3 (PIC).
FT CHAIN 496 795 COAT PROTEIN VP1 (PID).
FT CHAIN 796 984 CORE PROTEIN P2A.
FT CHAIN 985 1091 CORE PROTEIN P2B.
FT CHAIN 1092 1426 CORE PROTEIN P2C.
FT CHAIN 1427 1498 PROBABLE PROTEIN 3A.
FT CHAIN 1499 1521 PROBABLE PROTEIN 3B.
FT CHAIN 1522 1741 PROBABLE PROTEIN 3C.
FT CHAIN 1742 2230 RNA-DIRECTED POLYMERASE 3D.
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;

```

```

Query Match 94.2%; Score 97; DB 1; Length 2230;
Best Local Similarity 90.0%; Pred. No. 6.5e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPRK 20
DB 814 ORLKVAQEELSNEVLPPPRK 833

RESULT 11
POLG_HPAVT STANDARD; PRT; 839 AA.
AC P31788;
DT 01-JUL-1993 (Rel. 25, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN (CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEIN
DE P2A) (FRAGMENT).
OS Simian hepatitis A virus (strain CY-145).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=31707;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91311421; PubMed=1649902;
RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;
RT "Sequence analysis of a new hepatitis A virus naturally infecting
RT cynomolgus macaques (Macaca fascicularis).";
RL J. Gen. Virol. 72:1685-1689(1991).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS.
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59286; AAA5473.1; -.
DR PIR; J01180; GNNYS2.
KW Polyprotein; Coat protein; Core protein.
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
FT CHAIN 492 ? COAT PROTEIN VP1 (PID).
FT CHAIN ? >839 CORE PROTEIN P2A.
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93825 MW; 2CACC4BD1E192DBC CRC64;

Query Match 91.3%; Score 94; DB 1; Length 839;
Best Local Similarity 85.0%; Pred. No. 6.3e-07;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPRK 20
DB 809 ORLKVAQEELSNEVLPPPRK 828

RESULT 12
RPB2_CAPEL STANDARD; PRT; 1193 AA.
ID RPB2_CAPEL
AC Q10578;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II SECOND LARGEST SUBUNIT (EC 2.7.7.6)
DE (RNA POLYMERASE I SUBUNIT 2).
GN C26E6.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton L.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 200-1058 FROM N.A.
RX MEDLINE-95041334; PubMed=7953533;
RA Siodow A., Thomas W.K.;
RT "A molecular evolutionary framework for eukaryotic model organisms.";
RL Curr. Biol. 4:596-603(1994).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14
CC DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS THE SECOND LARGEST
CC COMPONENT OF RNA POLYMERASE II.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U13875; AAA21158.1; -  
 DR EMBL; U10333; AAA50224.1; -  
 DR WormPep; C26E6.4; CE01162.  
 DR InterPro; IP0001572; -  
 DR Pfam; PF00562; RNA\_pol\_B; 1.  
 DR PROSITE; PS01166; RNA-directed RNA polymerase; Transcription; Zinc;  
 KW Zinc-finger; Nuclear protein. C4-TYPE (POTENTIAL).  
 FT ZN\_FING 1125 1146  
 SQ SEQUENCE 1193 AA; 134904 MW; B8A85E74E9CC7EBE CRC64;

Query Match 50.5%; Score 52; DB 1; Length 1193;  
 Best Local Similarity 62.5%; Pred. No. 3.4;  
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLP 16  
 II:IIII:IIII  
 DB 340 ORIKYAREILQKLELP 355

RESULT 13  
 HSLU\_VIBCH STANDARD; PRT; 443 AA.  
 AC Q9KNO7:  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ATP-DEPENDENT HSL PROTEASE ATP-BINDING SUBUNIT HSLU.  
 GN HSLU OR VC2674.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae";  
 RL Nature 406:477-483(2000).  
 CC -!- FUNCTION: CHAPERONE SUBUNIT OF A PROTEASOME-LIKE DEGRADATION  
 CC COMPLEX (BY SIMILARITY).  
 CC -!- SUBUNIT: INTERACTS WITH HSLU (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CLPX CHAPERONE FAMILY. HSLU SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE004333; AAF95815.1; -  
 DR TIGR; VC2674; -  
 KW Chaperone; ATP-binding.  
 FT NP\_BIND 57 64 ATP (POTENTIAL).  
 SQ SEQUENCE 443 AA; 49900 MW; DAE13B82FAB6A38F CRC64;

Query Match 48.1%; Score 49.5; DB 1; Length 443;  
 Best Local Similarity 41.7%; Pred. No. 2.7;  
 Matches 10; Conservative 6; Mismatches 3; Indels 5; Gaps 1;

QY 1 ORLKVAQEELSNE-----VLPPPR 19  
 :::::IIII:IIII  
 DB 117 EKVKFRAEELAEERVDLPPPR 140

RESULT 14  
 RPB2\_HUMAN STANDARD; PRT; 1174 AA.  
 AC P30876;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II 140 KDA POLYPEPTIDE (EC 2.7.7.6)  
 DE (RNA POLYMERASE II SUBUNIT 2) (RPB2).  
 GN POLR2B.  
 GN Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=92389336; PubMed=1518060;  
 RA Acker J., Wintzerith M., Vigneron M., Keding C.;  
 RT "Primary structure of the second largest subunit of human RNA  
 polymerase II (or B).";  
 RL J. Mol. Biol. 226:1295-1299(1992).  
 RN [2]  
 RP SEQUENCE OF 699-1174 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97264341; PubMed=9110174;  
 RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,  
 RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;  
 RT "Large-scale concatenation cDNA sequencing.";  
 RL Genome Res. 7:353-358(1997).

CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X63563; CAA45124.1; -  
 DR EMBL; AF055028; AAC09367.1; -  
 DR PIR; S18986; S18986.  
 DR PIR; S28976; S28976.  
 DR MIM; 180661; -  
 DR InterPro; IPR001572; -  
 DR Pfam; PF00562; RNA\_pol\_B; 1.  
 DR PROSITE; PS01166; RNA\_POL\_BETA; 1.  
 KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;  
 KW Zinc-finger; Nuclear protein. C4-TYPE (POTENTIAL).  
 FT ZN\_FING 1119 1140  
 SQ SEQUENCE 1174 AA; 133896 MW; 32BEDF7F95E4DE10 CRC64;

Query Match 47.68; Score 49; DB 1; Length 1174;  
 Best Local Similarity 56.24; Pred. No. 9.8;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVL 16  
 :||:||||:| | | | |  
 DB 334 KRKYAKEVLQKEMLP 349

## RESULT 15

ID RPB2\_DROME STANDARD; PRT; 1176 AA.  
 AC P08266; Q95027; Q04155; Q9VEM7;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DE 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II 140 KDA POLYPEPTIDE (BC 2.7.7.6)  
 DE (RNA POLYMERASE II SUBUNIT 2).  
 GN RPII140 OR CG3180.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE OF 54-1176 FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=88011299; PubMed=3116266;  
 RA Falkenburg D., Dworniczak B., Faust D.M., Bautz E.K.F.;  
 RT "RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit  
 to the beta subunit of Escherichia coli RNA polymerase.";  
 RL J. Mol. Biol. 195:929-937(1987).  
 RN [2]  
 RP SEQUENCE OF 1-69 FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=91276237; PubMed=1905256;  
 RA Sitzler S., Oldenburg I., Peterson G., Bautz E.K.F.;  
 RT "Analysis of the promoter region of the housekeeping gene DmRPI40 by  
 sequence comparison of Drosophila melanogaster and Drosophila  
 virilis.";  
 RL Gene 100:155-162(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Dew C., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M.V., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 SUBSTRATES.  
 CC -I- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 RNA(N).  
 CC -I- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14  
 DIFFERENT POLYPEPTIDES. RNA POLYMERASE II CONSISTS OF 10 DIFFERENT  
 SUBUNITS. THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA  
 POLYMERASE II.  
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -I- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 III FOR 5S AND TRNA GENES.  
 CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X05709; CAA29180.2; -;  
 DR EMBL: M62972; AAA28476.1; -;  
 DR EMBL: AE003703; AAF55024.1; -;  
 DR PIR: A27826; A27826;  
 DR FlyBase: FBgn003276; RpiI140.  
 DR InterPro: IPR001572; -;  
 DR PROSITE: PS01166; RNA\_POL\_BETA; 1.  
 DR TRANSFERASE: DNA-directed RNA polymerase; Transcription; Zinc;  
 KW Zinc-finger; Metal-binding; Nuclear protein.  
 FT ZN\_FING 1121 1142 C4-TYPE (POTENTIAL).  
 FT CONFLICT 72 72 A -> R (IN REF. 1).  
 FT CONFLICT 666 667 ID -> MY (IN REF. 1).  
 SQ SEQUENCE 1176 AA; 134042 MW; 224821B335BED7F0 CRC64;

Query Match 46.68; Score 48; DB 1; Length 1176;  
 Best Local Similarity 56.24; Pred. No. 14;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ORLKYAOEELSNEVL 16  
 :||:||||:| | | | |  
 DB 335 KRKYAKEVLQKEMLP 350

Search completed: April 24, 2001, 16:33:58  
 Job time: 554 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:33:07 ; Search time 113.97 Seconds  
(without alignments)  
20.568 Million cell updates/sec

Title: US-09-171-432A-41  
Perfect score: 103  
Sequence: 1 ORLYAQEELSNEVLPPPRK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_15:  
1: sp-archaea:  
2: sp-bacteria:  
3: sp-fungi:  
4: sp-human:  
5: sp-invertebrate:  
6: sp-mammal:  
7: sp-mhc:  
8: sp-organelle:  
9: sp-phage:  
10: sp-plant:  
11: sp-rodent:  
12: sp-unclassified:  
13: sp-vertebrate:  
14: sp-virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	97	14	Q9QCT2 hepatitis a
2	103	100.0	97	14	Q9QCT1 hepatitis a
3	103	100.0	97	14	Q9QCT0 hepatitis a
4	103	100.0	97	14	Q9QCS9 hepatitis a
5	103	100.0	97	14	Q9QCS8 hepatitis a
6	103	100.0	97	14	Q9QCS7 hepatitis a
7	103	100.0	97	14	Q9QCS6 hepatitis a
8	103	100.0	97	14	Q9QCS5 hepatitis a
9	103	100.0	97	14	Q9QCS4 hepatitis a
10	103	100.0	97	14	Q9QCS3 hepatitis a
11	103	100.0	97	14	Q9QCS2 hepatitis a
12	103	100.0	97	14	Q9QCS1 hepatitis a
13	103	100.0	97	14	Q9QCS0 hepatitis a
14	103	100.0	97	14	Q9QCR9 hepatitis a
15	103	100.0	97	14	Q9QCR8 hepatitis a
16	103	100.0	97	14	Q9QCR7 hepatitis a
17	103	100.0	97	14	Q9QCR6 hepatitis a
18	103	100.0	97	14	Q9QCR5 hepatitis a
19	103	100.0	97	14	Q9QCR4 hepatitis a

20	103	100.0	97	14	Q9QCR3	Q9qcr3 hepatitis a
21	103	100.0	97	14	Q9QCR2	Q9qcr2 hepatitis a
22	103	100.0	97	14	Q9QCR1	Q9qcr1 hepatitis a
23	103	100.0	97	14	Q9QCR0	Q9qcr0 hepatitis a
24	103	100.0	97	14	Q9QCR9	Q9qcr9 hepatitis a
25	103	100.0	97	14	Q9QCR8	Q9qcr8 hepatitis a
26	103	100.0	97	14	Q9QCR7	Q9qcr7 hepatitis a
27	103	100.0	97	14	Q9QCR6	Q9qcr6 hepatitis a
28	103	100.0	97	14	Q9QCR5	Q9qcr5 hepatitis a
29	103	100.0	115	14	Q71976	Q71976 hepatitis a
30	103	100.0	116	14	Q71975	Q71975 hepatitis a
31	103	100.0	116	14	Q71977	Q71977 hepatitis a
32	103	100.0	116	14	Q71978	Q71978 hepatitis a
33	103	100.0	116	14	Q9W7X8	Q9w7x8 hepatitis a
34	103	100.0	116	14	Q9W7X7	Q9w7x7 hepatitis a
35	103	100.0	116	14	Q9W7T4	Q9w7t4 hepatitis a
36	103	100.0	116	14	Q9W7T0	Q9w7t0 hepatitis a
37	103	100.0	116	14	Q9W7S7	Q9w7s7 hepatitis a
38	103	100.0	116	14	Q92941	Q92941 hepatitis a
39	103	100.0	184	14	Q87092	Q87092 simian hepa
40	103	100.0	1124	14	Q84780	Q84780 hepatitis a
41	103	100.0	1161	14	Q05794	Q05794 hepatitis a
42	103	100.0	2216	14	Q9WMA2	Q9wma2 hepatitis a
43	103	100.0	2218	14	Q67824	Q67824 hepatitis a
44	103	100.0	2218	14	Q67817	Q67817 hepatitis a
45	103	100.0	2227	14	Q67825	Q67825 hepatitis a

## ALIGNMENTS

RESULT 1						
Q9QCT2		PRELIMINARY:				
ID	Q9QCT2		PRT:	97	AA.	
AC	Q9QCT2:					
DT	01-MAY-2000 (TRENBLrel. 13, Created)					
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)					
DT	01-MAY-2000 (TRENBLrel. 13, Last annotation update)					
DE	CAPSID PROTEIN (FRAGMENT).					
OS	Hepatitis A virus.					
OC	Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;					
OC	Hepatovirus.					
OX	NCBI_TaxID=12092;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=MIB36;					
RA	Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;					
RT	"A Multistate, Foodborne Outbreak of Hepatitis A.;"					
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AF199044; AAF08935.1; -					
FT	NON_TER 1					
FT	NON_TER 97					
SQ	SEQUENCE 97 AA; 11282 MW; 5ECB8F22812A62B9 CRC64;					
Query Match	100.0%;	Score	103;	DB	14;	Length 97;
Best Local Similarity	100.0%;	Pred. No.	1.1e-08;			
Matches	20;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
QY	1 ORLYAQEELSNEVLPPPRK 20					
DB	78 ORLYAQEELSNEVLPPPRK 97					
RESULT 2						
Q9QCT1		PRELIMINARY:				
ID	Q9QCT1		PRT:	97	AA.	
AC	Q9QCT1:					
DT	01-MAY-2000 (TRENBLrel. 13, Created)					
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)					
DT	01-MAY-2000 (TRENBLrel. 13, Last annotation update)					
DE	CAPSID PROTEIN (FRAGMENT).					
OS	Hepatitis A virus.					

```
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MIB35;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199045; AAF08936.1; -.
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11264 MW; 57C216228E259246 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 3
Q9QCT0
ID Q9QCT0 PRELIMINARY; PRT; 97 AA.
AC Q9QCT0;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MIB4032;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199046; AAF08937.1; -.
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11282 MW; 5ECB8F22812A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 4
Q9QCS9
ID Q9QCS9 PRELIMINARY; PRT; 97 AA.
AC Q9QCS9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MIB212;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
```

```
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199047; AAF08938.1; -.
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 5
Q9QCS8
ID Q9QCS8 PRELIMINARY; PRT; 97 AA.
AC Q9QCS8;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MIB25;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199048; AAF08939.1; -.
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 6
Q9QCS7
ID Q9QCS7 PRELIMINARY; PRT; 97 AA.
AC Q9QCS7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BUTTE5;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199049; AAF08940.1; -.
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11296 MW; 5ED5401E6D7B27AC CRC64;
```

Query Match 100.0%; Score 103; DB 14; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 78 ORLKVAQEELSNEVLPPPRK 97

RESULT 7  
Q9QCS6 PRELIMINARY; PRT; 97 AA.  
ID Q9QCS6;  
AC Q9QCS6;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE CAPSID PROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SLC92;  
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF199050; AAF08941.1; -;  
FT NON\_TER 1  
FT NON\_TER 97  
SQ SEQUENCE 97 AA; 11296 MW; 5ED5401E6D7B27AC CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 78 ORLKVAQEELSNEVLPPPRK 97

RESULT 8  
Q9QCS5 PRELIMINARY; PRT; 97 AA.  
ID Q9QCS5;  
AC Q9QCS5;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE CAPSID PROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SLC778;  
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF199051; AAF08942.1; -;  
FT NON\_TER 1  
FT NON\_TER 97  
SQ SEQUENCE 97 AA; 11268 MW; 4897242E6D7B27AC CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||

DB 78 ORLKVAQEELSNEVLPPPRK 97

RESULT 9  
Q9QCS4 PRELIMINARY; PRT; 97 AA.  
ID Q9QCS4;  
AC Q9QCS4;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE CAPSID PROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WI4005;  
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF199052; AAF08943.1; -;  
FT NON\_TER 1  
FT NON\_TER 97  
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 78 ORLKVAQEELSNEVLPPPRK 97

RESULT 10  
Q9QCS3 PRELIMINARY; PRT; 97 AA.  
ID Q9QCS3;  
AC Q9QCS3;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE CAPSID PROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BCM202;  
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF199053; AAF08944.1; -;  
FT NON\_TER 1  
FT NON\_TER 97  
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKVAQEELSNEVLPPPRK 20  
|||||  
DB 78 ORLKVAQEELSNEVLPPPRK 97

RESULT 11  
Q9QCS2 PRELIMINARY; PRT; 97 AA.  
ID Q9QCS2;  
AC Q9QCS2;

```
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AZA7;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199054; AAF08945.1; -
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 12
Q9QCS1
ID Q9QCS1 PRELIMINARY; PRT; 97 AA.
AC Q9QCS1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ME2001;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199055; AAF08946.1; -
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 13
Q9QCS0
ID Q9QCS0 PRELIMINARY; PRT; 97 AA.
AC Q9QCS0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
```

```
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LA5015;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199056; AAF08947.1; -
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 14
Q9QCR9
ID Q9QCR9 PRELIMINARY; PRT; 97 AA.
AC Q9QCR9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BUTTEL;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF199057; AAF08948.1; -
FT NON_TER 1
FT NON_TER 97
SQ SEQUENCE 97 AA; 11310 MW; 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20
Db 78 QRLKYAQEELSNEVLPPPRK 97

RESULT 15
Q9QCR8
ID Q9QCR8 PRELIMINARY; PRT; 97 AA.
AC Q9QCR8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CAPSID PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AZA3;
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
```



DR EMBL; AF199058; AAF08949.1; -  
 FT NON\_TER 1  
 FT NON\_TER 97  
 SQ SEQUENCE 97 AA: 11310 MW: 5ED5401E6C2A62B9 CRC64;

Query Match 100.0%; Score 103; DB 14; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 Db 78 QRLKYAQEELSNEVLPPPRK 97

Search completed: April 24, 2001, 16:33:08  
 Job time: 569 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:23:17 ; Search time 121.9 Seconds  
(without alignments)  
9.379 Million cell updates/sec

Title: US-09-171-432A-42

Perfect score: 101

Sequence: 1 VLPPPRMKGLFSQAKISLF 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	101	100.0	20	18 W42925	Immunogenic Hepati
2	101	100.0	366	6 P50230	Sequence of hepati
3	101	100.0	854	6 P50287	Sequence encoded b
4	101	100.0	993	6 P50116	Sequence of Hepati
5	101	100.0	993	6 P50231	Sequence encoded b
6	101	100.0	1077	20 W95559	A partial hepatic
7	101	100.0	1091	14 R32426	Translated from 5'
8	101	100.0	1091	14 R05697	Attenuated hepatic
9	101	100.0	2227	11 W34074	Hepatitis A virus
10	101	100.0	2227	18 B18607	Amino acid sequenc
11	101	100.0	2227	21 B18608	Amino acid sequenc

12	101	100.0	2227	21 B18609	Amino acid sequenc
13	96	95.0	2227	7 P60066	Sequence of viral
14	84	83.2	839	12 R15629	Capsid region of c
15	65	64.4	25	18 W42930	Immunogenic Hepati
16	52.5	52.0	1025	16 R86001	Human dihydropyrim
17	52.5	52.0	1025	17 R91420	Human dihydropyrim
18	52.5	52.0	1025	20 W93361	Human DPD protein.
19	47	46.5	407	21 G06085	Arabidopsis thalia
20	45.5	45.0	1025	16 R86000	Bovine dihydropyri
21	45	44.6	23	12 R15517	BMP-8 peptide. Bo
22	45	44.6	119	21 G05136	Arabidopsis thalia
23	45	44.6	119	21 G49796	Arabidopsis thalia
24	45	44.6	132	21 G05135	Arabidopsis thalia
25	45	44.6	132	21 G49795	Arabidopsis thalia
26	45	44.6	150	21 G05134	Arabidopsis thalia
27	45	44.6	150	21 G36068	Zea mays protein f
28	45	44.6	150	21 G49794	Arabidopsis thalia
29	45	44.6	178	21 G05133	Arabidopsis thalia
30	45	44.6	178	21 G36067	Arabidopsis thalia
31	45	44.6	178	21 G49793	Arabidopsis thalia
32	44	43.6	187	21 B32794	Eucalyptus grandis
33	42	41.6	20	18 W42926	Immunogenic Hepati
34	42	41.6	71	21 G26746	Zea mays protein f
35	42	41.6	87	21 G26745	Zea mays protein f
36	42	41.6	113	21 B40306	Human OREF ORF70 p
37	42	41.6	131	21 G26744	Zea mays protein f
38	42	41.6	150	21 G41992	Arabidopsis thalia
39	42	41.6	178	21 G41991	Arabidopsis thalia
40	42	41.6	193	21 G41990	Arabidopsis thalia
41	41	40.6	108	17 R90801	Reverse-Frame HGV
42	41	40.6	108	19 W80170	Clone G15 amino ac
43	41	40.6	108	19 W76076	HGV clone GE15 pro
44	41	40.6	108	20 W92777	US5856134 Seq ID 1
45	41	40.6	108	20 W89486	Hepatitis G virus

ALIGNMENTS

RESULT 1  
W42925  
ID W42925 standard; peptide; 20 AA.  
XX  
AC W42925;  
XX  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1318.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW immune response; antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
PN W09740147-AL.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US06891.  
XX  
PR 19-APR-1996; 96US-0015644.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an  
PT immune response to HAV in a mammal or to detect the presence of  
PT antibodies against HAV in a mammal  
XX

PS Claim 18; Page 112; 140pp; English.

XX Peptides W42922-30 are immunogenic peptides corresponding to immunogenic  
CC epitopes of the Hepatitis A virus (HAV). The peptides are substantially  
CC similar to a portion of the amino acid sequence of the P2A protein of HAV  
CC corresponding to amino acids 792-980. The present peptide is derived  
CC from amino acids 823-842, and has a reactivity of 31.3% with acute sera.  
CC Compositions containing the peptides can be used to induce an immune  
CC response to HAV in a mammal. The peptides can also be used to detect the  
CC presence of antibodies against HAV in mammalian serum. The peptides can  
CC also be used to make an antibody against HAV by administering the peptide  
CC to a mammal.

XX Sequence 20 AA;

Query Match 100.0%; Score 101; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 1 vlppprkmgkglfsqakisl 20

RESULT 2

P50230  
ID P50230 standard; Protein; 366 AA.

XX AC P50230;

XX DT 28-NOV-1991 (first entry)

XX DE Sequence of hepatitis A virus (HAV) surface protein (VP-1).

XX KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;

XX KW diagnostic assay.

XX OS Hepatitis A virus.

XX PN EP138704-A.

XX PD 24-APR-1985.

XX PF 09-OCT-1984; 84EP-0402025.

XX PR 02-MAR-1984; 84US-0585942.

XX PR 14-OCT-1983; 83US-0541836.

XX PA (MERI ) MERCK & CO INC.

XX PI Hughes JV, Scolnick EM, Tomassini JE;

XX DR WPI; 1985-100818/17.

XX DR N-PSDB; N50274.

XX PT New hepatitis A virus surface protein - useful for binding to  
XX PT neutralising antibodies to the virus

XX PS Claim 21; Page 46-48; 49pp; English.

XX CC VPI is isolated by solubilisation of the intact virus in an aq.

XX CC anionic surfactant and a reducing agent. The viral proteins are sepd.

XX CC and the protein of molecular wt. 33000 daltons is sepd.

XX SQ Sequence 366 AA;

Query Match 100.0%; Score 101; DB 6; Length 366;

Best Local Similarity 100.0%; Pred. No. 2.3e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

DB 332 vlppprkmgkglfsqakisl 351

RESULT 3

ID P50287 standard; Protein; 854 AA.

XX AC P50287;

XX DT 30-NOV-1991 (first entry)

XX DE Sequence encoded by hepatitis A virus (HAV) cDNA from near the  
XX DE genome 5' terminus to the end of the area corresponding to the  
XX DE capsid protein region of poliovirus RNA.

XX KW Hepatitis A virus assay; antigen; antibody.

XX OS Hepatitis A virus.

XX PN W08501517-A.

XX PD 11-APR-1985.

XX PF 27-SEP-1984; 84WO-US01552.

XX PR 30-SEP-1983; 83US-0537911.

XX PA (MASI ) MASSACHUSETTS INST TECH.

XX PI Ticehurst JR, Baltimore D, Feinstone SM, Purcell RH;  
XX PI Racaniello VR;

XX DR WPI; 1985-098846/16.

XX DR N-PSDB; N50330.

XX PT New hepatitis A virus cDNA - useful in assays for the virus and  
XX PT for prodn. of the viral antigen and antibodies to it

XX PS Example; Fig 7; 60pp; English.

XX CC The inventors claim HAV cDNA and a method for producing it, whereby  
XX CC large amts. can be obtd. economically. The cDNA is useful in the  
XX CC assay for detection of HAV quickly and easily and with high  
XX CC sensitivity and specificity. The HAV cDNA is also used in the prodn.  
XX CC of HAV antigen or antibodies to it. The antibodies may be monoclonal.

XX SQ Sequence 854 AA;

Query Match 100.0%; Score 101; DB 6; Length 854;

Best Local Similarity 100.0%; Pred. No. 5.5e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

DB 823 vlppprkmgkglfsqakisl 842

RESULT 4

ID P50116 standard; Protein; 993 AA.

XX AC P50116;

XX DT 30-SEP-1991 (first entry)

XX DE Sequence of Hepatitis A virus (HAV) immunogenic peptides  
XX DE VP-1, VP-2, VP-3 and VP-4.

XX KW Antigenic protein; immunogen; vaccine.

XX OS Hepatitis A virus (strain CR326).

```

PR   02-MAR-1984;    84US-0585942.
XX   14-OCT-1983;    83US-0541836.
XX   (MERI ) MERCK & CO INC.
XX
XX   Hughes JV, Scolnick EM, Tomassini JE;
PI
XX
XX   WPI; 1985-100818/17.
XX   N-PSDB; N5027A.
XX
XX   New hepatitis A virus surface protein - useful for binding to
PT   neutralising antibodies to the virus
XX
XX   Disclosure: Page 17-23; 49pp; English.
XX
XX   VP1 is isolated by solubilisation of the intact virus in an aq.
CC   anionic surfactant and a reducing agent. The viral proteins are sepd
CC   and the protein of molecular wt. 33000 daltons is sepo.
XX
XX   Sequence      993 AA;
SQ
Query Match              100.0%; Score 101; DB 6; Length 993;
Best Local Similarity    100.0%; Pred.No. 6.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY       1 VLPPrKMkMGLFSQAkISLF 20
          |||||
DB       959 vlppprkmkglfsqakisl f 978
          |||||
RESULT        6
W95559
ID     W95559 standard; Protein; 1077 AA.
XX
XX     W95559;
XX
XX   28-APR-1999 (first entry)
XX
XX   A partial hepatitis A virus (HAV) protein..
XX
XX   Hepatitis A virus protein; HAV; P2 region;
KW   cell-culture-adapted HAV strain; infection; accelerated growth.
KW
XX   Hepatitis A virus.
OS
XX
XX   US5849562-A.
XX
XX   15-DEC-1998.
XX
XX   06-JUN-1995; 95US-0468926.
XX
XX   06-NOV-1991; 91US-0788262.
PR   30-SEP-1983; 83US-0537911.
PR   27-SEP-1984; 84US-0654942.
PR   06-OCT-1988; 88US-0256135.
PR   06-JUN-1995; 95US-0468926.
XX
XX   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX   Emerson SU, Purcell RH;
XX
XX   WPI; 1999-094412/08.
DR   N-PSDB; X01006.
XX
XX   Chimeric hepatitis A virus strains - with P2 region from
PT   cell-culture-adapted strain in wild-type genome
XX
XX   Disclosure; Fig 7A-L; 36pp; English.
XX
XX   The present sequence represents a partial hepatitis A virus (HAV)
CC   protein. The specification describes a DNA construct consisting
CC   of a wild-type HAV genome in which the P2 region is replaced by the

```

CC P2 region from a cell-culture-adapted HAV strain. The construct is  
 CC used to demonstrate that mutations in the P2 region of a  
 CC cell-culture-adapted HAV strain are sufficient for establishment of  
 CC infection and accelerated growth in cell culture.

XX  
 SQ Sequence 1077 AA;

Query Match 100.0%; Score 101; DB 20; Length 1077;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20  
 |||||  
 Db 1046 vlppprkmgkglfsqakisl 1065

RESULT 7

R32426  
 ID R32426 standard; Protein; 1091 AA.

XX  
 AC R32426;

XX  
 DT 10-JUN-1993 (first entry)

XX Translated from 5' region of Hepatitis A Virus genomic clone.

XX HAV HM-175; chronic liver disease; picornavirus.

XX Hepatitis A Virus.

XX Key Location/Qualifiers  
 XX 238..1091  
 FT Region /label= ORF

FT /note= "second putative initiation codon at  
 FT position 240"

FT Region 1..711

FT /note= "X's correspond to nonsense codons,  
 FT i.e. this region is not an ORF"

XX US7788262-A.

XX 15-DEC-1992.

XX 30-SEP-1993; 83US-0536911.

XX 27-SEP-1984; 84US-0654942.

XX 06-OCT-1988; 88US-0256135.

XX 30-SEP-1983; 83US-0536911.

XX 06-NOV-1991; 91US-0788262.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.

XX Baltimore D, Feinstein SM;

XX Purcell RH, Racanelli VR, Ticehurst JR;

XX WPI; 1993-067429/08.

XX N-PSDB; Q36934.

XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.  
 XX of antigen and antibodies

XX Disclosure; Fig 7; 65pp; English.

XX HAV virion RNA was extracted from the livers of marmosets which had  
 CC been inoculated with HAV (the HAV had previously been passaged twice  
 CC in marmosets). The RNA was used to prepare ds cDNA clones by  
 CC standard methods. Clones contg. inserts which hybridised to RNA from  
 CC HAV-infected African Green Monkey Kidney cells were selected for  
 CC further analysis. A 7.4kb restriction map (about 99% of the HAV  
 CC genome) was constructed from 5 overlapping inserts. The sequence of  
 CC the first 3.3kb (approx.) from the 5'-terminus was determined. An  
 CC amino acid sequence was decoded from the entire clone and an open

CC reading frame was identified starting at position 238. A comparison  
 CC of the predicted HAV amino acid sequences with the known capsid  
 CC protein sequences of other picornaviruses (poliovirus, foot and  
 CC mouth disease virus and encephalomyelitis virus) revealed areas of  
 CC local homology.

XX  
 SQ Sequence 1091 AA;

Query Match 100.0%; Score 101; DB 14; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 7e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20  
 |||||  
 Db 1060 vlppprkmgkglfsqakisl 1079

RESULT 8

R05697  
 ID R05697 standard; protein; 2227 AA.

XX  
 AC R05697;

XX 15-AUG-1990 (first entry)

XX Attenuated hepatitis A virus.

XX Hepatitis A virus; vaccine; attenuated.

XX Hepatitis A virus, strain HM-175.

XX Key Location/Qualifiers

FT Region 1..23

FT /label=VP4 = 1A

FT Region 24..245

FT /label=VP2 = 1B

FT Region 246..491

FT /label=VP3 = 1C

FT Region 492..791

FT /label=VP1 = 1D

FT Region 792..980

FT /label=2A

FT Region 981..1087

FT /label=2B

FT Region 1088..1422

FT /label=2C

FT Region 1423..1496

FT /label=3A

FT Region 1497..1519

FT /label=3B = VPg

FT Region 1520..1738

FT /label=3C

FT Region 1739..2227

FT /label=3D

XX US4894228-A.

XX 16-JAN-1990.

XX 12-JUL-1988; 88US-0217824.

XX 12-JUL-1988; 88US-0217824.

XX 12-JUL-1988; 88US-0652967.

XX (USSH ) US DEPT HEALTH & HUMAN.

XX Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;

XX Daemer RJ, Gust ID;

XX WPI; 1990-075557/10.

XX N-PSDB; Q03512.

PT Vaccine against hepatitis A virus infection - comprises novel  
 PT attenuated hepatitis A virus strain.  
 XX  
 PS Claim 1; Fig 1; 18pp; English.  
 XX  
 CC The attenuated HAV is useful for inducing protective immunity against  
 CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
 CC several nucleotide changes distributed throughout the genome, is  
 CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
 CC suitable for use as an HAV vaccine. It is noted that not all the changes  
 CC are necessary for attenuation and use as a vaccine.  
 XX  
 SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 11; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 vlppprkmgkglfsqakisl 842

RESULT 9  
 ID W34074 standard; Protein; 2227 AA.  
 AC W34074;  
 DT 27-APR-1998 (first entry)  
 XX  
 XX Hepatitis A virus HM-175 protein sequence.  
 DE  
 XX HAV; attenuation: 2C protein; 2C chimeric hepatitis A virus;  
 KW infection; vaccine.  
 XX  
 OS Hepatitis A virus HM-175.  
 XX

Key Location/Qualifiers  
 FH 1..23  
 FT Protein /label= VP4  
 FT 24..245  
 FT Protein /label= VP2  
 FT 246..491  
 FT Protein /label= VP3  
 FT 492..791  
 FT Protein /label= VP1  
 FT 792..980  
 FT Protein /label= 2A  
 FT 981..1087  
 FT Protein /label= 2B  
 FT 1088..1422  
 FT Protein /label= 2C  
 FT 1423..1496  
 FT Protein /label= 3A  
 FT 1497..1519  
 FT Protein /label= 3B  
 FT 1520..1738  
 FT Protein /label= 3C  
 FT 1739..2227  
 FT Protein /label= 3D  
 XX  
 XX WO9740166-A2.  
 XX  
 XX 30-OCT-1997.  
 XX  
 XX 18-APR-1997; 97WO-US06506.  
 XX  
 XX 19-APR-1996; 96US-0015642.  
 XX  
 XX (USSH ) US SEC DEPT HEALTH.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA

XX Emerson SU, Purcell RH, Raychaudhuri G;  
 PI WPI; 1997-535850/49.  
 XX N-PSDB; T93023.  
 DR  
 XX Human attenuated HAV genome containing simian HAV 2C gene - useful  
 PT as vaccines against HAV infection  
 XX  
 PS Disclosure; Fig 13A-D; 66pp; English.  
 XX  
 CC This protein sequence is encoded by the human hepatitis A virus  
 CC (HAV) HM-175 wild-type genome (see T93023). Attenuated strain  
 CC HAV/7 is obtained by passage of HM-175 in African Green Monkey  
 CC kidney cells. A claimed DNA construct (I) comprises a genome of  
 CC HAV, where the genome is a human attenuated HAV genome in which a  
 CC region of the 2C gene has been replaced by a corresponding region  
 CC from a 2C gene of a simian AGM-27 HAV genome (see T93024). The  
 CC region of the 2C gene from AGM-27 contained in the construct  
 CC preferably encodes amino acids 120-328 of the 2C protein, amino  
 CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA  
 CC transcript of (I); (2) a cell transfected with (I) or the RNA  
 CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the  
 CC HAV of (3); and (5) a host cell containing the HAV of (3). (I) or  
 CC its RNA transcript, can be used as a vaccine for preventing HAV in  
 CC a mammal. (I) or the RNA transcript can also be used to stimulate  
 CC the production of protective antibodies in the mammal.  
 XX  
 SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 18; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 vlppprkmgkglfsqakisl 842

RESULT 10  
 ID B18607 standard; Protein; 2227 AA.  
 XX  
 AC B18607;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.  
 XX  
 KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN US6113912-A.  
 XX  
 PD 05-SEP-2000.  
 XX  
 PF 07-JUN-1995; 95US-0475886.  
 XX  
 PR 18-SEP-1992; 92US-0947338.  
 PR 17-SEP-1993; 93WO-US08610.  
 PR 10-MAR-1995; 95US-0397232.  
 XX  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 PI WPI; 2000-586464/55.  
 XX N-PSDB; A75476.  
 DR  
 XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type -  
 XX  
 XX Disclosure; Fig 6A-K; 72pp; English.

CC The present sequence is derived from a wild type hepatitis A virus  
 CC (HAV) strain HM-174. The sequence is modified to produce HAV which  
 CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
 CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
 CC attenuation. It is useful as a live vaccine for prophylaxis of  
 CC hepatitis A in humans and other primates.

XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VLPPrKMKGLFSQAKISLF 20

Db 823 vlpprkmkgifsqakisl 842

RESULT 11

BI8608

ID BI8608 standard; Protein; 2227 AA.

AC BI8608;

XX 15-JAN-2001 (first entry)

DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

KW P-35 virus.

XX Hepatitis A virus.

OS US6113912-A.

PN 05-SEP-2000.

XX 07-JUN-1995; 95US-0475886.

XX 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93WO-US08610.

PR 10-MAR-1995; 95US-0397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

XX WPI; 2000-586464/55.

DR N-PSDB; A75477.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type -

XX Disclosure; Columns 67-78; 72pp; English.

XX The present sequence is derived from passage 35 of a wild type

CC hepatitis A virus (HAV) strain HM-174. The resulting virus is

CC designated P-35 virus. The sequence is modified to produce HAV which

CC are adapted to growth in the human fibroblast-like cell line MRC-5.

CC The HAV is able to propagate in MRC-5 cells and retain appropriate

CC attenuation. It is useful as a live vaccine for prophylaxis of

CC hepatitis A in humans and other primates.

XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VLPPrKMKGLFSQAKISLF 20

Db 823 vlpprkmkgifsqakisl 842

RESULT 12

BI8609

ID BI8609 standard; Protein; 2227 AA.

XX AC BI8609;

XX 15-JAN-2001 (first entry)

DE Amino acid sequence of live attenuated Hepatitis A virus 4380.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

KW HAV 4380.

XX Hepatitis A virus.

OS US6113912-A.

PN 05-SEP-2000.

XX 07-JUN-1995; 95US-0475886.

XX 18-SEP-1992; 92US-0947338.

PR 17-SEP-1993; 93WO-US08610.

PR 10-MAR-1995; 95US-0397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

XX WPI; 2000-586464/55.

DR N-PSDB; A75478.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type -

XX Disclosure; Columns 93-104; 72pp; English.

XX The present sequence is derived from a live attenuated hepatitis A

CC virus (HAV) of the invention, designated HAV 4380. The sequence is

CC produced by modifying wild type HAV strain HM-174. The HAV of the

CC invention are adapted to growth in the human fibroblast-like cell

CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain

CC appropriate attenuation. It is useful as a live vaccine for prophylaxis

CC of hepatitis A in humans and other primates.

XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VLPPrKMKGLFSQAKISLF 20

Db 823 vlpprkmkgifsqakisl 842

RESULT 13

P60066

ID P60066 standard; Protein; 2227 AA.

XX AC P60066;

XX 26-JUN-1991 (first entry)



XX Sequence of viral L434 polypeptide encoded by the complete  
DE nucleotide sequence of the HAV genome.

XX  
KW Diagnosis; vaccine; passive immunotherapy.

XX Hepatitis A virus.

XX Key Location/Qualifiers

FT Region 1..245

FT Region /label= P1.1A

FT Region 246..491

FT Region /label= 1B

FT Region 492..836

FT Region /label= 1C

FT Region 837..980

FT Region /label= P2.2A

FT Region 981..1076

FT Region /label= 2B

FT Region 1077..1422

FT Region /label= 2C

FT Region 1423..1484

FT Region /label= P3.3A

FT Region 1485..1507

FT Region /label= 3B

FT Region 1508..1678

FT Region /label= 3C

FT Region 1679..2227

FT Region /label= 3D

XX EP199480-A.

XX 29-OCT-1986.

XX 03-APR-1986; 86EP-0302465.

XX 03-APR-1985; 85US-0719329.

XX (CHIR-) CHIRON CORP.

XX Dina D, Potter SJ, Vannest GA, Caput D;

XX WPI; 1986-286213/44.

XX N-PSDB; N60080.

XX Hepatitis A virus nucleotide sequence and polypeptide - and use

XX in prodn. of vaccines and diagnostic probes

XX Claim 5; Fig 1; 18pp; English.

XX N60080 and oligonucleotide fragments are useful in detection of

XX hepatitis A virus; transformed hosts may be used for expression of

XX polypeptides and fragments useful in vaccines without risk of

XX infection by the virus or in prodn. of particles which are capable

XX of inducing immunocompetent B cells for passive immunotherapy. Pref.

XX epitope is derived from AAS 445-657 or 792-848 of the HAV

XX polypeptide sequence (P60066).

XX Sequence 2227 AA;

XX Query Match 95.0%; Score 96; DB 7; Length 2227;

XX Best Local Similarity 95.0%; Pred. No. 9e-07;

XX Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

DB 823 vlppprkrmnglfsqakislf 842

RESULT 14

ID R15629 standard; Protein; 839 AA.

XX R15629;

XX 17-MAR-1992 (first entry)

XX Capsid region of cyno-HAV isolate CY-145.

XX Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.

XX Cynomolgus monkey hepatitis A virus, isolate CY-145.

XX Key Location/Qualifiers

FT Protein 1..245

FT Protein /label= VP0

FT Protein 246..491

FT Protein /label= VP3

FT Protein 492..791

FT Protein /label= VP1

FT Protein 792

FT Protein /label= P2

FT Protein /note= "incomplete"

FT Cleavage-site 245..246

FT Cleavage-site 491..492

FT Cleavage-site 791..792

FT Active-site 315

FT Active-site 593

XX US7678828-A.

XX 12-NOV-1991.

XX 03-APR-1991; 91US-0678828.

XX 03-APR-1991; 91US-0678828.

XX (USSH) US DEPT HEALTH & HUMAN.

XX Hainan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;

XX WPI; 1991-376737/51.

XX N-PSDB; Q15180.

XX Hepatitis A virus isolates and DNA - used to prepare vaccines for

XX preventing hepatitis A virus infection.

XX Disclosure; Fig 3; 23pp; English.

XX The sequence was deduced from the nucleotide sequence obt'd. by PCR

XX amplification of cyno-HAV viral RNA obt'd. from the stool of a

XX cynomolgus monkey with serologically and histologically confirmed

XX spontaneous hepatitis A. The sequence differs from the human HAV

XX isolate HM175 (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci.

XX USA 84, 2497-2501), mainly in the VP3 and VP1 proteins. The Gln-Val

XX pair at the VP3-VP1 cleavage site in the human isolate is replaced

XX by a Gln-Thr pair in the cyno-HAV. The other two cleavage sites are

XX the same. Two residues have been identified as part of the immuno-

XX dominant region (see feature table) and are different to those in

XX the same position in human HAV. The protein and peptides derived

XX from it can be used in the prepn. of vaccines for the prevention of

XX HAV infection.

XX See also R15056.

XX Sequence 839 AA;

XX Query Match 83.2%; Score 84; DB 12; Length 839;

XX Best Local Similarity 83.3%; Pred. No. 2.7e-05;

XX Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKIS 18

DB 822 ilppprkmgllfsqakis 839

```

RESULT 15
W42930
ID W42930 standard; peptide; 25 AA.
XX
AC W42930;
XX
DT 28-APR-1998 (first entry)
XX
DE Immunogenic Hepatitis A virus peptide YK-1665.
XX
KW Immunogenic peptide; immunogenic epitope; P2A protein;
KW immune response; antibody.
XX
OS Synthetic.
OS Hepatitis A virus.
XX
PN W09740147-A1.
XX
PD 30-OCT-1997.
XX
PF 18-APR-1997; 97WO-US06891.
XX
PR 19-APR-1996; 96US-0015644.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fields HA, Khudyakov YE;
XX
DR WPI; 1997-535831/49.
XX
PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an
PT immune response to HAV in a mammal or to detect the presence of
PT antibodies against HAV in a mammal
XX
PS Claim 18; Page 112; 140pp; English.
XX
CC Peptides W42922-30 are immunogenic peptides corresponding to immunogenic
CC epitopes of the Hepatitis A virus (HAV). The peptides are substantially
CC similar to a portion of the amino acid sequence of the P2A protein of HAV
CC corresponding to amino acids 792-980. Compositions containing the
CC peptides can be used to induce an immune response to HAV in a mammal.
CC The peptides can also be used to detect the presence of antibodies
CC against HAV in mammalian serum. The peptides can also be used to make an
CC antibody against HAV by administering the peptide to a mammal.
XX
SQ Sequence 25 AA;

```

Query Match 64.4%; Score 65; DB 18; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.00085;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 VLPPPRKMKGLF 12
   | | | | | | | | | |
Db 14 vlppprkmgkglf 25

```

Search completed: April 24, 2001, 16:23:22  
 Job time: 140 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:35:00 ; Search time 62.1 Seconds  
(without alignments)  
6.187 Million cell updates/sec

Title: US-09-171-432a-42  
Perfect score: 101  
Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pcp:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pcp:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pcp:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pcp:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pcp:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	1091	6 5516630-2	Patent No. 5516630
2	101	100.0	2227	4 US-08-475-886-2	Sequence 2, Appli
3	101	100.0	2227	4 US-08-475-886-4	Sequence 4, Appli
4	101	100.0	2227	4 US-08-475-886-6	Sequence 6, Appli
5	101	100.0	2227	4 US-08-397-232-2	Sequence 2, Appli
6	101	100.0	2227	4 US-08-397-232-4	Sequence 4, Appli
7	84	83.2	839	1 US-08-087-016-2	Sequence 2, Appli
8	52.5	52.0	1025	2 US-08-304-309-2	Sequence 2, Appli
9	52.5	52.0	1025	3 US-08-991-942-2	Sequence 2, Appli
10	52.5	52.0	1025	5 PCT-US95-04567-4	Sequence 4, Appli
11	45.5	45.0	1025	5 PCT-US95-04567-2	Sequence 2, Appli
12	45	44.6	23	1 US-07-800-364B-4	Sequence 4, Appli
13	45	44.6	23	5 PCT-US91-03388-4	Sequence 4, Appli
14	41	40.6	108	1 US-08-466-033-106	Sequence 106, App
15	41	40.6	108	2 US-08-444-733-106	Sequence 106, App
16	41	40.6	108	2 US-08-464-134-106	Sequence 106, App
17	41	40.6	108	2 US-08-461-361-106	Sequence 106, App
18	41	40.6	108	2 US-08-485-910-106	Sequence 106, App
19	41	40.6	108	5 PCT-US95-06266-89	Sequence 89, Appl
20	41	40.6	616	1 US-08-638-911A-35	Sequence 35, Appl
21	41	40.6	853	1 US-08-638-911A-25	Sequence 25, Appl
22	41	40.6	853	1 US-08-638-911A-27	Sequence 27, Appl
23	41	40.6	853	1 US-08-638-911A-29	Sequence 29, Appl
24	41	40.6	853	1 US-08-638-911A-31	Sequence 31, Appl
25	41	40.6	2873	1 US-08-466-033-15	Sequence 15, Appl
26	41	40.6	2873	1 US-08-638-911A-2	Sequence 2, Appli
27	41	40.6	2873	2 US-08-444-733-15	Sequence 15, Appli

28	41	40.6	2873	2 US-08-464-134-15	Sequence 15, Appl
29	41	40.6	2873	2 US-08-461-361-15	Sequence 15, Appl
30	41	40.6	2873	2 US-08-485-910-15	Sequence 15, Appl
31	41	40.6	2873	5 PCT-US95-06266-15	Sequence 15, Appl
32	41	40.6	2910	2 US-08-466-033-183	Sequence 183, App
33	41	40.6	2910	2 US-08-444-733-183	Sequence 183, App
34	41	40.6	2910	2 US-08-464-134-183	Sequence 183, App
35	41	40.6	2910	2 US-08-461-361-183	Sequence 183, App
36	41	40.6	2910	2 US-08-485-910-183	Sequence 183, App
37	41	40.6	2910	5 PCT-US95-06266-157	Sequence 157, App
38	40	39.6	34	2 US-08-751-767A-38	Sequence 38, Appl
39	40	39.6	166	1 US-08-129-456A-11	Sequence 11, Appl
40	40	39.6	310	1 US-08-129-456A-36	Sequence 36, Appl
41	40	39.6	310	1 US-08-129-456A-37	Sequence 37, Appl
42	40	39.6	310	2 US-08-705-868-3	Sequence 3, Appli
43	40	39.6	310	3 US-09-123-615-3	Sequence 3, Appli
44	40	39.6	371	2 US-08-442-809A-76	Sequence 76, Appl
45	40	39.6	391	2 US-08-644-034A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
5516630-2  
; Patent No. 5516630  
; APPLICANT: TICEHURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE, STEPHEN M.; PURCELL, ROBERT H.; RACANELLO, VINCENT R.; BAROUDY, BAHIGE M.  
; TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/788,262  
; FILING DATE: 06-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 256,135  
; FILING DATE: 06-OCT-1988  
; APPLICATION NUMBER: 654,942  
; FILING DATE: 27-SEP-1984  
; APPLICATION NUMBER: 537,911  
; FILING DATE: 30-SEP-1983  
; SEQ ID NO: 2:  
; LENGTH: 1091  
5516630-2

Query Match 100.0% Score 101; DB 6; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 5.5e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||  
DB 1060 VLPPPRKMKGLFSQAKISLF 1079

RESULT 2  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20
Db 823 VLPPrKMKGLFSQAKISLF 842

RESULT 3
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 2026426US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20
Db 823 VLPPrKMKGLFSQAKISLF 842

RESULT 4
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 2026426US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20
Db 823 VLPPrKMKGLFSQAKISLF 842

RESULT 5
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 2026426US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20
Db 823 VLPPrKMKGLFSQAKISLF 842

RESULT 6
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 2026426US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4
```

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||  
DB 823 VLPPPRKMKGLFSQAKISLF 842

## RESULT 7

US-08-087-016-2  
; Sequence 2, Application US/08087016  
; Patent No. 5430135  
; GENERAL INFORMATION:  
; APPLICANT: NAINAN, OMANA V.  
; APPLICANT: MARGOLIS, HAROLD S.  
; APPLICANT: ROBERTSON, BETTY H.  
; APPLICANT: BRINTON, MARGO H.  
; APPLICANT: EBERT, JAMES W.  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,016  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,828  
; FILING DATE: 03-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 839 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 83.2%; Score 84; DB 1; Length 839;  
Best Local Similarity 83.3%; Pred. No. 3e-06;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKIS 18  
:|||||:|||||:|||||  
DB 822 ILPPPRKMKGLFSQSKIS 839

## RESULT 8

US-08-304-309-2  
; Sequence 2, Application US/08304309  
; Patent No. 5856454  
; GENERAL INFORMATION:

; APPLICANT: GONZALEZ, Frank J.  
; APPLICANT: FERNANDEZ-SALGUERO, Pedro  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF cDNA FOR HUMAN  
; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/304,309  
; FILING DATE: 09-SEP-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Timothy L.  
; REGISTRATION NUMBER: 35,367  
; REFERENCE/DOCKET NUMBER: 15280-210  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1025 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-304-309-2

Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.78;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPRKMKGLFSQAKISLF 20  
||||| : |||:|  
DB 176 LPPEKMESEAYS-AKIALF 193

## RESULT 9

US-08-991-942-2  
; Sequence 2, Application US/08991942  
; Patent No. 6015673  
; GENERAL INFORMATION:  
; APPLICANT: GONZALEZ, Frank J.  
; APPLICANT: FERNANDEZ-SALGUERO, Pedro  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF cDNA FOR HUMAN  
; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/991,942  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/304,309  
FILING DATE: 09-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 15280-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-991-942-2

Query Match 52.0%; Score 52.5; DB 3; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.78;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Qy 2 LPPPRKMKGLFSQAKISLF 20  
|||||| :| |||:|  
Db 176 LPPPEKMSEAYS-AKIALF 193

RESULT 10  
PCT-US95-04567-4  
Sequence 4, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: PatentIn Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-4

Query Match 52.0%; Score 52.5; DB 5; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 0.78;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Qy 2 LPPPRKMKGLFSQAKISLF 20  
|||||| :| |||:|  
Db 176 LPPPEKMSEAYS-AKIALF 193

RESULT 11  
PCT-US95-04567-2  
Sequence 2, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: PatentIn Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-2

Query Match 45.0%; Score 45.5; DB 5; Length 1025;  
Best Local Similarity 61.1%; Pred. No. 12;  
Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Qy 2 LPPPRKMKGLFSQAKISLF 19  
|||||| :| |||:|  
Db 176 LPPPEKMSEAYS-AKIALF 192

RESULT 12  
US-07-800-364B-4  
Sequence 4, Application US/07800364B  
Patent No. 568678  
GENERAL INFORMATION:  
APPLICANT: Hewick, Rodney M.  
APPLICANT: Wang, Jack H.  
APPLICANT: Wozney, John M.  
APPLICANT: Celeste, Anthony J.

TITLE OF INVENTION: Bone and Cartilage Inductive Proteins  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/800,364B  
FILING DATE: 26-NOV-1991  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kapinos, Ellen J.  
REGISTRATION NUMBER: 32,245  
REFERENCE/DOCKET NUMBER: GI 5182A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-876-1170  
TELEFAX: 617-876-5851  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Bos taurus  
TISSUE TYPE: Bone  
US-07-800-364B-4

Query Match 44.6%; Score 45; DB 1; Length 23;  
Best Local Similarity 63.6%; Pred. No. 0.23;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPPPRKMKGLF 12  
Db 4 LPPPNKLGIF 14

RESULT 13  
PCT-US91-03388-4  
Sequence 4, Application PC/TUS9103388  
GENERAL INFORMATION:  
APPLICANT: Hewick, Rodney M.  
APPLICANT: Wang, Jack H.  
TITLE OF INVENTION: Bone and Cartilage Inductive Proteins  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/03388  
FILING DATE: 19910515  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:

NAME: Kapinos, Ellen J.  
REGISTRATION NUMBER: 32,245  
REFERENCE/DOCKET NUMBER: G15182X-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-876-1170  
TELEFAX: 617-876-5851  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Bos taurus  
TISSUE TYPE: Bone  
PCT-US91-03388-4

Query Match 44.6%; Score 45; DB 5; Length 23;  
Best Local Similarity 63.6%; Pred. No. 0.23;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPPPRKMKGLF 12  
Db 4 LPPPNKLGIF 14

RESULT 14  
US-08-466-033-106  
Sequence 106, Application US/08466033  
Patent No. 5766840  
GENERAL INFORMATION:  
APPLICANT: Kim, Jungsuh P.  
APPLICANT: Wages, John  
APPLICANT: Young, Lavonne M.  
APPLICANT: Fry, Kirk E.  
APPLICANT: Linnen, Jeffrey M.  
TITLE OF INVENTION: Hepatitis G Virus and Molecular  
Cloning Thereof  
NUMBER OF SEQUENCES: 277  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,033  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/389,886  
FILING DATE: 15-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/357,509  
FILING DATE: 16-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,729  
FILING DATE: 26-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,271  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/285,558  
FILING DATE: 03-AUG-1994

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/285,543  
;; FILING DATE: 03-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/246,985  
;; FILING DATE: 20-MAY-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fabian, Gary R.  
;; REGISTRATION NUMBER: 33,875  
;; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0960  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 106:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 108 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-466-033-106

Query Match 40.6%; Score 41; DB 1; Length 108;  
Best Local Similarity 46.7%; Pred. No. 5.9;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKI 17  
||| |||:|:|:

Db 57 PPPVPKGVFKKAPL 71

## RESULT 15

US-08-444-733-106  
; Sequence 106, Application US/08444733  
; Patent No. 5624507

## ; GENERAL INFORMATION:

;; APPLICANT: Kim, Jungshuh P.  
;; APPLICANT: Wages, John  
;; APPLICANT: Young, LaVonne M.  
;; APPLICANT: Fry, Kirk E.  
;; APPLICANT: Linnen, Jeffrey M.  
;; TITLE OF INVENTION: Hepatitis G Virus and Molecular  
;; TITLE OF INVENTION: Cloning Thereof  
;; NUMBER OF SEQUENCES: 277  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Dehlinger & Associates  
;; STREET: 350 Cambridge Ave., Suite 250  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94306

## ; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/444,733  
;; FILING DATE:

## ; CLASSIFICATION:

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/389,886  
;; FILING DATE: 15-FEB-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/357,509  
;; FILING DATE: 16-DEC-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/329,729  
;; FILING DATE: 26-OCT-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/344,271  
;; FILING DATE: 23-NOV-1994  
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/285,558  
;; FILING DATE: 03-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/285,543  
;; FILING DATE: 03-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/246,985  
;; FILING DATE: 20-MAY-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fabian, Gary R.  
;; REGISTRATION NUMBER: 33,875  
;; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 324-0960  
;; TELEFAX: (415) 324-0960  
;; INFORMATION FOR SEQ ID NO: 106:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 108 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-444-733-106

Query Match 40.6%; Score 41; DB 2; Length 108;  
Best Local Similarity 46.7%; Pred. No. 5.9;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKI 17  
||| |||:|:|:

Db 57 PPPVPKGVFKKAPL 71

Search completed: April 24, 2001, 16:35:02  
Job time: 533 sec





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:25:51 ; Search time 73.14 Seconds  
(without alignments)  
18.792 Million cell updates/sec

Title: US-09-171-432A-42

Perfect score: 101  
Sequence: 1 VLPPPRMKGLFSQAKISLF 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 58722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_67:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	101	100.0	852	1 GNNYHA	genome polyprotein
2	101	100.0	1358	2 A03905	genome polyprotein
3	101	100.0	2227	1 GNNYHM	genome polyprotein
4	101	100.0	2227	1 GNNYHR	genome polyprotein
5	101	100.0	2227	1 GNNYMK	genome polyprotein
6	96	95.0	2230	1 GNNYSA	genome polyprotein
7	95	94.1	2227	1 GNNYHB	genome polyprotein
8	84	83.2	839	1 GNNYS2	genome polyprotein
9	75	74.3	341	2 S04137	genome polyprotein
10	52.5	52.0	1025	2 A54718	dihydropyrimidine
11	44	43.6	340	1 WMBES7	ribonucleoside-dip
12	44	43.6	343	2 S74937	hypothetical prote
13	44	43.6	369	2 S77299	C4-dicarboxylase-b
14	44	43.6	775	2 T48957	hypothetical prote
15	43	42.6	309	2 D75008	methanol dehydrog
16	43	42.6	379	2 T49919	hypothetical prote
17	43	42.6	389	2 S41748	heat shock protein
18	42.5	42.1	105	2 B72598	hypothetical prote
19	42	41.6	109	2 S69307	probable membrane
20	42	41.6	165	2 S02053	hypothetical prote
21	42	41.6	178	2 T02335	ribosomal protein
22	42	41.6	834	2 F82673	hypothetical prote
23	41.5	41.1	304	2 H82984	hypothetical prote
24	41	40.6	83	2 T17809	hypothetical prote
25	41	40.6	159	2 C72210	conserved hypothet
26	41	40.6	229	2 T19153	hypothetical prote
27	41	40.6	237	2 C84065	transcription regu
28	41	40.6	622	2 T37257	hypothetical prote
29	41	40.6	897	2 T06540	polyribonucleotide

30	41	40.6	1040	2 A34695	axonal glycoprotei
31	41	40.6	1234	2 T30254	jumonji protein -
32	40.5	40.1	924	2 T25007	hypothetical prote
33	40.5	40.1	932	2 T25008	hypothetical prote
34	40	39.6	177	2 C64133	mercuric resistanc
35	40	39.6	238	2 C83240	probable transcrip
36	40	39.6	259	2 F64532	conserved hypothet
37	40	39.6	272	2 T49070	probable heat shoc
38	40	39.6	292	2 S23239	hypothetical prote
39	40	39.6	310	2 A55053	endothelial monocy
40	40	39.6	310	2 B55053	endothelial monocy
41	40	39.6	371	2 I45089	thyroid transcript
42	40	39.6	372	2 S53724	thyroid-specific
43	40	39.6	372	2 S12002	thyroid nuclear fa
44	40	39.6	380	2 T20269	hypothetical prote
45	40	39.6	401	2 G02321	thyroid transcript

ALIGNMENTS

RESULT 1

GNNYHA

genome polyprotein - human hepatitis A virus (strain CR326) (fragment)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A

C:Species: human hepatitis A virus

A:Note: host Homo sapiens (man)

C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999

C:Accession: A03904

R:Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra,

J. Virol. 54, 247-255, 1985

A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.

A:Reference number: A03904; MUID:85185648

A:Accession: A03904

A:Molecule type: genomic RNA

A:Residues: 1-852 <LIN>

A:Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; polyprotein

F:1-245/Product: coat protein 1A #status predicted <C1A>

F:246-491/Product: coat protein 1B #status predicted <C1B>

F:492-836/Product: coat protein 1C #status predicted <C1C>

F:837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 101; DB 1; Length 852;

Best Local Similarity 100.0%; Pred. No. 1e-08; 0; Mismatches 0; Gaps 0;

Matches 20; Conservative 0; Indels 0; Gaps 0;

QY 1 VLPPPRMKGLFSQAKISLF 20

|||||

Db 823 VLPPPRMKGLFSQAKISLF 842

RESULT 2

A03905

genome polyprotein (version 2) - human hepatitis A virus (fragments)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; prote

C:Species: human hepatitis A virus

C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996

C:Accession: A03905

R:Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feins

Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985

A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and R

A:Reference number: A03905; MUID:85166289

A:Accession: A03905

A:Molecule type: genomic RNA

A:Residues: 1-1358 <BAR>

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; polyprotein

F:1-245/Product: coat protein 1A #status predicted <C1A>

F:246-491/Product: coat protein 1B #status predicted <C1B>

F:492-836/Product: coat protein 1C #status predicted <C1C>

C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase

A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:g222597; PIDN:BAA00766.1; PID:g222598  
R:Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure and reference sequence  
A:Reference number: JQ1080; MUID:91311420  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhikov, V.B.  
Submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:g61971; PIDN:CAA33490.1; PID:g930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhikov, V.B.  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian hepatitis A virus  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: coat protein 2A #status predicted <C2A>  
F:985-1091/Product: coat protein 2B #status predicted <C2B>  
F:1092-1426/Product: coat protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 95.0%; Score 96; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 2.1e-07;  
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
:|||||:|||||:|||||  
Db 827 ILPPRKVKGLFSQAKISLF 846

RESULT 7  
GNNYHB  
genome polyprotein - human hepatitis A virus (strain MBB)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein VPg; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J50303  
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wissel, T.; Kiehn, R.; Wimmer, E.; Deinhardt, Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (isolated from a patient with acute hepatitis A)  
A:Reference number: J50303; MUID:88045071  
A:Accession: J50303  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <PAU>  
A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; hydrophobic protein  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP1>  
F:837-980/Product: core protein 2A #status predicted <P2A>

F:981-1108/Product: core protein 2B #status predicted <P2B>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1496/Product: protein 3A #status predicted <P3A>  
F:1497-1519/Product: genome-linked protein VPg #status predicted <VPG>  
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 94.1%; Score 95; DB 1; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 3e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
:|||||:|||||:|||||  
Db 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 8  
GNNYS2  
genome polyprotein - simian hepatitis A virus (strain CY-145) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat protein 1C; protein 3A  
C:Species: simian hepatitis A virus  
A:Note: host Macaca fascicularis (cynomolgus macaque)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
C:Accession: JQ1180  
R:Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A:Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus monkeys  
A:Reference number: JQ1180; MUID:91311421  
A:Accession: JQ1180  
A:Molecule type: genomic RNA  
A:Residues: 1-839 <NAI>  
A:Cross-references: GB:M59286; NID:g329599; PIDN:AAA45473.1; PID:g555083  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; glycoprotein; polyprotein  
F:1-23/Product: coat protein 1A #status predicted <VP0>  
F:24-245/Product: coat protein 1B #status predicted <VP3>  
F:246-491/Product: coat protein 1C #status predicted <VP1>  
F:492-839/Product: core protein 2A (fragment) #status predicted <P2P>  
F:261,312,728,756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 83.2%; Score 84; DB 1; Length 839;  
Best Local Similarity 83.3%; Pred. No. 6.9e-06;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKIS 18  
:|||||:|||||:|||||  
Db 822 ILPPPRKMKGLFSQAKIS 839

RESULT 9  
S04137  
genome polyprotein - human hepatitis A virus (strain LCDC-1) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: S04137  
R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus  
A:Reference number: S04137; MUID:89263805  
A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>  
A:Cross-references: EMBL:X14666; NID:g62301; PIDN:CAA32794.1; PID:g4377576  
C:Genetics:  
A:Gene: VP1  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; polyprotein  
F:2-340/Product: coat protein 1D (VP1) #status predicted <NAT>

Query Match 74.3%; Score 75; DB 2; Length 341;

Best Local Similarity 93.3%; Pred. No. 8.1e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQA 15  
|||||  
Db 327 VLPPrKMKGLFSQS 341

RESULT 10  
A54718  
dihydropyrimidine dehydrogenase (NADP+) (EC 1.3.1.2) - human  
C:Species: Homo sapiens (man)  
C:Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 17-Mar-2000  
C:Accession: A54718  
R:Yokota, H.; Fernandez-Salguero, P.; Furiya, H.; Lin, K.; McBride, O.W.; Podschun, B.;  
J. Biol. Chem. 269, 23192-23196, 1994  
A:Title: cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, a  
A:Reference number: A54718; MUID:94365020  
A:Accession: A54718  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1025 <YOK>  
A:Cross-references: GB:009178; NID:g558304; PID:g558305  
A:Note: authors translated the codon GGC for residue 748 as Asp, and TTA for residue 101  
C:Genetics:  
A:Gene: GDB:DPYD  
A:Cross-references: GDB:364102; OMIM:274270  
A:Map position: lp22-lp22  
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] h  
C:Keywords: 4Fe-4S; flavoprotein; homodimer; iron-sulfur protein; NADP; oxidoreductase  
F:946-1005/domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 1.5;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPrKMKGLFSQAKISLF 20  
|||||  
Db 176 LPPPrKMKSEAYS-AKIALF 193

RESULT 11  
WNB57  
ribonucleoside-diphosphate reductase (EC 1.17.4.1) small chain - human herpesvirus 1 (st  
N:Alternate names: ribonucleotide reductase small chain  
C:Species: human herpesvirus 1  
C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jun-2000  
C:Accession: D30088  
R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perr  
J. Gen. Virol. 69, 1531-1574, 1988  
A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim  
A:Reference number: A30083; MUID:88274327  
A:Accession: D30088  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-340 <MCG>  
A:Cross-references: GB:X14112; NID:gl944536; PIDN:CAA32303.1; PID:g59540; GB:D00317  
C:Genetics:  
A:Gene: UL40  
C:Function:  
A:Description: catalyzes the reduction of ribonucleoside diphosphate to deoxyribonucleos  
C:Superfamily: herpesvirus ribonucleoside-diphosphate reductase small chain  
C:Keywords: deoxyribonucleotide biosynthesis; early protein; iron; metalloprotein; oxid  
F:94,124,127,221,224/Binding site: 2Fe-O cluster (Asp, Glu, His, Glu, Glu, His) #sta  
F:131/Active site: Tyr (stable tyrosyl radical) #status predicted

Query Match 43.6%; Score 44; DB 1; Length 340;  
Best Local Similarity 61.5%; Pred. No. 12;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPrKMKGLFSQA 15

|||||  
Db 242 PPrDPrVYGLFRQA 254

RESULT 12  
S74937  
hypothetical protein slr0688 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S74937  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
S.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S74937  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-343 <KAN>  
A:Cross-references: EMBL:D90902; GB:AB001339; NID:gl652027; PIDN:BAAL6977.1; PID:d101  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 43.6%; Score 44; DB 2; Length 343;  
Best Local Similarity 47.1%; Pred. No. 12;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPrKMKGLFSQAKISL 19  
|||||  
Db 240 PPrECKGLFRGAETI 256

RESULT 13  
S77299  
C4-dicarboxylase-binding protein - Synechocystis sp. (strain PCC 6803)  
N:Alternate names: protein sl11314  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S77299  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
S.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S77299  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-369 <KAN>  
A:Cross-references: EMBL:D90907; GB:AB001339; NID:gl652618; PIDN:BAAL7633.1; PID:d101  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Gene: dctP

Query Match 43.6%; Score 44; DB 2; Length 369;  
Best Local Similarity 50.0%; Pred. No. 13;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQAKISLF 20  
|||||  
Db 38 ILPLPGKYGYYSQAKIRAF 57

RESULT 14  
T48957  
hypothetical protein T1583.170 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T48957

R: Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unsel, M.; Mewes, H.W.; Rudd, S.;  
submitted to the Protein Sequence Database, April 2000

A: Reference number: Z25009  
A: Accession: T48957  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-775 <J0R>  
A: Cross-references: EMBL:AL163975; GSPDB:GN00061; ATSP:T15B3.170  
A: Experimental source: cultivar Columbia; BAC clone T15B3  
C: Genetics:  
A: Gene: ATSP:T15B3.170  
A: Map position: 3  
A: Introns: 33/3; 135/3; 298/1; 376/2; 478/3; 551/3

Query Match 43.6%; Score 44; DB 2; Length 775;  
Best Local Similarity 52.9%; Pred. No. 29;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
QY 3 PPPKMKGLFSQAKISL 19  
||| ||| ||  
Db 749 PPPRNKHELNFQCHTSL 765

RESULT 15  
D75008  
methanol dehydrogenase regulatory protein (moxr-2) PAB1296 - Pyrococcus abyssi (strain Q  
C: Species: Pyrococcus abyssi  
C: Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C: Accession: D75008  
R: anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A: Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
A: Reference number: A75001  
A: Accession: D75008  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-309 <KAW>  
A: Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50506.1; PID:g545902  
A: Experimental source: strain Orsay  
C: Genetics:  
A: Gene: PAB1296  
C: Superfamily: methanol dehydrogenase regulatory protein

Query Match 42.6%; Score 43; DB 2; Length 309;  
Best Local Similarity 81.8%; Pred. No. 16;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 6 RKMKGFLFSQAK 16  
|| ||||| ||  
Db 185 RKDKGLFSEAK 195

Search completed: April 24, 2001, 16:25:54  
Job time: 288 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:33:58 ; Search time 38.74 Seconds  
(without alignments)  
17.685 Million cell updates/sec

Title: US-09-171-432A-42  
Perfect score: 101  
Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	852	1 POLG_HPAVC	P06442 hepatitis a
2	101	100.0	2227	1 POLG_HPAVH	P08617 hepatitis a
3	101	100.0	2227	1 POLG_HPAVL	P06441 hepatitis a
4	96	95.0	2226	1 POLG_HPAV2	P26580 hepatitis a
5	96	95.0	2226	1 POLG_HPAV4	P26581 hepatitis a
6	96	95.0	2226	1 POLG_HPAV8	P26582 hepatitis a
7	96	95.0	2230	1 POLG_HPAVS	P14553 simian hepa
8	95	94.1	2227	1 POLG_HPAVM	P13901 hepatitis a
9	84	83.2	839	1 POLG_HPAVT	P31788 simian hepa
10	75	74.3	341	1 POLG_HPAV1	P13672 hepatitis a
11	52.5	52.0	1025	1 DPYD_HUMAN	Q12882 homo sapien
12	45.5	45.0	1025	1 DPYD_BOVIN	Q28007 bos taurus
13	44	43.6	340	1 RIR2_HSV11	P10224 herpes simp
14	44	43.6	340	1 RIR2_HSV1K	P06474 herpes simp
15	43	42.6	389	1 DNAJ_METHA	P5515 methanosarc
16	43	42.6	587	1 MTSJ_STRAL	Q53609 streptomyce
17	41	40.6	289	1 HK26_MOUSE	P43688 mus musculu
18	41	40.6	1040	1 AXOL_RAT	P22063 rattus norv
19	41	40.6	1234	1 JWI_MOUSE	Q62315 mus musculu
20	40	39.6	135	1 ZNTR_HAEIN	P45277 haemophilus
21	40	39.6	292	1 YOWI_CAEEL	P30647 caenorhabdi
22	40	39.6	371	1 TTF1_CANFA	P43698 canis famil
23	40	39.6	371	1 TTF1_HUMAN	P43699 homo sapien
24	40	39.6	372	1 TTF1_MOUSE	P50220 mus musculu
25	40	39.6	372	1 TTF1_RAT	P23441 rattus norv
26	40	39.6	448	1 DHE4_HELPY	P55990 helicobacte
27	40	39.6	523	1 YMR9_YEAST	Q05040 saccharomyc
28	40	39.6	1048	1 SRA4_RAT	Q63627 rattus norv
29	40	39.6	1157	1 SRA4_HUMAN	O95104 homo sapien
30	39.5	39.1	1040	1 MAN1_RAT	P21139 rattus norv
31	39	38.6	204	1 RS4_TREPA	O83328 treponema p
32	39	38.6	347	1 EI3A_SOYBN	Q03773 glycine max
33	39	38.6	427	1 TRB1_AERPE	Q09875 aeropyrum p

ALIGNMENTS

RESULT 1

POLG\_HPAVC STANDARD; PRT; 852 AA.  
AC P06442: Q83741; Q83742;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEIN P2A] (FRAGMENT).  
DE Hepatitis A virus (strain CR326).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID-12097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-85185648; PubMed-2985793;  
RA Linemeyer D.L.; Menke J.G.; Martin-Gallardo A.; Hughes J.V.;  
RA Young A.; Mitra S.W.;  
RT "Molecular Cloning and partial sequencing of hepatitis A viral cDNA.";  
RL J. Virol. 54:247-255(1985).  
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; M10033; AAA45470.1; -;  
DR PIR; A03904; GNNVHA  
KW Polypeptide; Coat protein; Core protein.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).  
FT CHAIN 837 >852 CORE PROTEIN P2A.  
FT NON\_TER 852 852  
SQ SEQUENCE 852 AA; 95363 MW; 73D3ED0AD532820E CRC64;

Query Match 100.0%; Score 101; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 4.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||  
DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 2

POLG\_HPAHV

ID AC P08617; P06443; STANDARD; PRT; 2227 AA.

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)].

DE Hepatitis A virus (strain HM-175).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus.

OX NCBI\_TaxID=12098;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WILD TYPE.

RX MEDLINE=87061253; PubMed=3023706;

RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A., Baroudy B.M.;

RT "Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other Picornaviruses.";

RT J. Virol. 61:50-59(1987).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATTENUATED;

RX MEDLINE=87175701; PubMed=3031686;

RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstein S.M., Purcell R.H.;

RT "Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA polymerase.";

RT Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).

RL -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.

CC -1- VP3, AND VP4.

CC -1- SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT SHOWN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M14114; AAA45475.1; -

DR EMBL; M14707; AAA45465.1; -

DR EMBL; M14707; AAA45466.1; ALT\_INIT.

DR EMBL; M16632; AAA45471.1; -

DR PIR; A25981; GNNYHM.

DR PIR; A25914; GNNYMK.

DR PIR; A03905; A03905.

DR MEROPS; C03.005; -

DR InterPro; IPR000605; -

DR Pfam; PF00680; RNA\_dep\_RNA\_pol.1.

DR PIR; A03910; RNA\_helicase.1.

KW Polyprotein; Coat protein; Core protein; Transferase;

KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.

FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).

FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).

FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).

FT CHAIN 492 836 COAT PROTEIN VP1 (PID).

FT CHAIN 837 980 CORE PROTEIN P2A.

FT CHAIN 981 1087 CORE PROTEIN P2B.

FT CHAIN 1088 1422 CORE PROTEIN P2C.

FT CHAIN 1423 1496 PROBABLE PROTEIN P3A.

FT CHAIN 1497 1519 PROBABLE PROTEIN P3B.

FT CHAIN 1520 1738 PROBABLE PROTEIN P3C.

FT CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.

FT VARIANT 77 77 K -> R (IN ATTENUATED STRAIN).

FT VARIANT 764 764 E -> V (IN ATTENUATED STRAIN).

FT VARIANT 821 821 N -> S (IN ATTENUATED STRAIN).

FT VARIANT 1052 1052 A -> V (IN ATTENUATED STRAIN).

FT VARIANT 1062 1062 G -> A (IN ATTENUATED STRAIN).

FT VARIANT 1118 1118 K -> M (IN ATTENUATED STRAIN).

FT VARIANT 1151 1151 E -> K (IN ATTENUATED STRAIN).

FT VARIANT 1163 1163 F -> S (IN ATTENUATED STRAIN).

FT VARIANT 1277 1277 V -> I (IN ATTENUATED STRAIN).

FT VARIANT 1500 1500 H -> Y (IN ATTENUATED STRAIN).

FT VARIANT 1805 1805 D -> N (IN ATTENUATED STRAIN).

FT VARIANT 1930 1930 S -> T (IN ATTENUATED STRAIN).

SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.3e-08; Mismatches 0; Indels 0; Caps 0;

Matches 20; Conservative 0;

QY 1 VLPPTKMKGLFSQAKISLF 20

DB 823 VLPPTKMKGLFSQAKISLF 842

RESULT 3

POLG\_HPAVL STANDARD; PRT; 2227 AA.

ID AC P06441;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)].

DE Hepatitis A virus (strain LA).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus.

OX NCBI\_TaxID=12099;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85190549; PubMed=2986127;

RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A., Merryweather J., van Nest G., Dina D.;

RT "Primary structure and gene organization of human hepatitis A virus.";

RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).

CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.

CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; K02990; AAA45472.1; -

DR PIR; A03903; GNNYHR.

DR MEROPS; C03.005; -

DR InterPro; IPR000605; -

```

DR  InterPro: IPR001205; RNA_dep_RNA_pol; 1.
DR  Pfam: PF00680; RNA_helicase; 1.
DR  Pfam: PF00910; RNA_helicase; 1.
KW  Polyprotein; Coat protein; Core protein; Thiol protease;
KW  RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT  CHAIN 1 23
FT  CHAIN 24 245
FT  CHAIN 246 491
FT  CHAIN 492 836
FT  CHAIN 837 980
FT  CHAIN 981 1076
FT  CHAIN 1077 1422
FT  CHAIN 1423 1484
FT  CHAIN 1485 1507
FT  CHAIN 1508 1678
FT  CHAIN 1679 2227
SQ  SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPrKKMKGLFSQAKISLF 20
    |||||
DB 823 VLPPrKKMKGLFSQAKISLF 842

RESULT 4
POLG_HPAV2 STANDARD; PRT; 2226 AA.
AC P26580;
DT 01-AUG-1992 (Rel. 23; Created)
DT 01-AUG-1992 (Rel. 23; Last sequence update)
DT 15-DEC-1998 (Rel. 37; Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 24a).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromeans T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M59810; AAA45468.1;
CC MEROPS; C03.005;
CC InterPro: IPR001205;
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Thiol protease;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT  CHAIN 1 23
FT  CHAIN 24 245
FT  CHAIN 246 491
FT  CHAIN 492 836
FT  CHAIN 837 980
FT  CHAIN 981 1076
FT  CHAIN 1077 1422
FT  CHAIN 1423 1484
FT  CHAIN 1485 1507
FT  CHAIN 1508 1678
FT  CHAIN 1679 2227
SQ  SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

```

```

FT  CHAIN 1 23
FT  CHAIN 24 245
FT  CHAIN 246 491
FT  CHAIN 492 836
FT  CHAIN 837 980
FT  CHAIN 981 1076
FT  CHAIN 1077 1422
FT  CHAIN 1423 1484
FT  CHAIN 1485 1507
FT  CHAIN 1508 1678
FT  CHAIN 1679 2227
SQ  SEQUENCE 2226 AA; 251152 MW; 6CD85A91D684E2BF CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2226;
Best Local Similarity 95.0%; Pred. No. 8.9e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPrKKMKGLFSQAKISLF 20
    |||||
DB 823 VLPPrKKMKGLFSQAKISLF 842

RESULT 5
POLG_HPAV4 STANDARD; PRT; 2226 AA.
AC P26581;
DT 01-AUG-1992 (Rel. 23; Created)
DT 01-AUG-1992 (Rel. 23; Last sequence update)
DT 15-DEC-1998 (Rel. 37; Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 43c).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12095;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromeans T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M59809; AAA45469.1;
CC MEROPS; C03.005;
CC InterPro: IPR001205;
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Thiol protease;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT  CHAIN 1 23
FT  CHAIN 24 245
FT  CHAIN 246 491
FT  CHAIN 492 836
FT  CHAIN 837 980
FT  CHAIN 981 1076
FT  CHAIN 1077 1422
FT  CHAIN 1423 1484
FT  CHAIN 1485 1507
FT  CHAIN 1508 1678
FT  CHAIN 1679 2227
SQ  SEQUENCE 2226 AA; 251152 MW; 6CD85A91D684E2BF CRC64;

```

```

FT CHAIN 901 1087 CORE PROTEIN P2B.
FT CHAIN 1088 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2226;
Best Local Similarity 95.0%; Pred. No. 8.9e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQAKISLF 20
Db 823 VLPPrKMKGLFSQAKISLF 842
|||||
RESULT 6
POLG_HPAV8 STANDARD; PRT; 2226 AA.
ID POLG_HPAV8
AC P26582;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Simian hepatitis A virus (strain 18f).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromeans T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065(1991).
RC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M59808; AAA5467.1; -
CC MEROPS; C03.005; -
CC InterPro; IPR000605; -
CC InterPro; IPR001205; -
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC Polyprotein; Coat protein; Core protein; Thiol protease.
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).
FT CHAIN 795 900 CORE PROTEIN P2A.
FT CHAIN 901 1087 CORE PROTEIN P2B.
FT CHAIN 1088 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.

FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2226;
Best Local Similarity 95.0%; Pred. No. 8.9e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQAKISLF 20
Db 823 VLPPrKMKGLFSQAKISLF 842
|||||
RESULT 7
POLG_HPAVS STANDARD; PRT; 2230 AA.
ID POLG_HPAVS
AC P14533;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Simian hepatitis A virus (strain AGM-27).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12102;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91311420; PubMed=1649901;
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,
RA Purcell R.H.;
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome
RT structure and growth in cell culture with other HAV strains."
RL J. Gen. Virol. 72:1677-1683(1991).
RC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00924; BAA00766.1; -
CC EMBL; X15461; CAA33490.1; -
CC PIR; A30470; GNNYSA.
CC PIR; S04885; S04885.
CC MEROPS; C03.005; -
CC InterPro; IPR000605; -
CC InterPro; IPR001205; -
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC Polyprotein; Coat protein; Core protein; Thiol protease.
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 27 COAT PROTEIN VP4 (PIA).
FT CHAIN 28 249 COAT PROTEIN VP2 (PIB).
FT CHAIN 250 495 COAT PROTEIN VP3 (PIC).
FT CHAIN 496 795 COAT PROTEIN VP1 (PID).
FT CHAIN 796 984 CORE PROTEIN P2A.

```

```

FT CHAIN 985 1091 CORE PROTEIN P2B.
FT CHAIN 1092 1426 CORE PROTEIN P2C.
FT CHAIN 1427 1498 PROBABLE PROTEIN 3A.
FT CHAIN 1499 1521 PROBABLE PROTEIN 3B.
FT CHAIN 1522 1741 PROBABLE PROTEIN 3C.
FT CHAIN 1742 2230 RNA-DIRECTED POLYMERASE 3D.
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324EIF19 CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2230;
Best Local Similarity 90.0%; Pred. No. 8.9e-08;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRMKGLFSQAKISLF 20
:|||||:|||||
DB 827 ILPPPRKVKGLFSQAKISLF 846

RESULT 8
POLG_HPVMV
ID POLG_HPVMV STANDARD; PRT: 2227 AA.
AC P1301; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;
AC Q81090; Q81091; Q81092; Q81093;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain MBB).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88045071; PubMed=2823500;
RA Paul A.V., Tada H., der Helm K., Wissel T., Kiehn R., Wimmer E.,
RT Deinhardt F.;
RT "The entire nucleotide sequence of the genome of human hepatitis A
RT virus (isolate MBB).";
RL Virus Res. 8:153-171(1987).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M20273; AAA45474.1;
CC PIR: JS0303; GNNYHB.
CC DR MEROPS: C03.005;
CC DR InterPro: IPR000605;
CC DR InterPro: IPR001205;
CC DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
CC DR Pfam: PF00910; RNA_helicase; 1.
CC DR Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).
FT CHAIN 837 980 CORE PROTEIN P2A.
FT CHAIN 981 1087 CORE PROTEIN P2B.
FT CHAIN 1088 1422 CORE PROTEIN P2C.

```

AC P13672;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP3; CORE PROTEIN  
DE P2A] (FRAGMENT).  
OS Hepatitis A virus (strain LCD-1).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12093;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89263805; PubMed=2542903;  
RA Andonov A.P., Lau P., Chaudhary R.;  
RT "Nucleotide sequence of the vp1 gene from a Chinese strain of  
RT hepatitis A virus (HAV).";  
RL Nucleic Acids Res. 17:3594-3594(1989).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X14666; CAA32794.1; -  
DR PIR; S04137; S04137.  
KW Polyprotein; Coat protein; Core protein.  
FT NON\_TER 1 1  
FT CHAIN <1 1 COAT PROTEIN VP3 (1C).  
FT CHAIN 2 340 COAT PROTEIN VP1 (1D).  
FT CHAIN 341 >341 CORE PROTEIN P2A.  
FT NON\_TER 341 341  
SQ SEQUENCE 341 AA; 38003 MW; 066918289BFF126D5 CRC64;  
  
Query Match 74.3%; Score 75; DB 1; Length 341;  
Best Local Similarity 93.3%; Pred. No. 3.8e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VLPPPRKMKGLFSQA 15  
|||||  
DB 327 VLPPPRKMKGLFSQS 341  
  
RESULT 11  
DPYD\_HUMAN  
ID DPYD\_HUMAN STANDARD; PRT; 1025 AA.  
AC Q12882; Q16694; Q16761;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+] PRECURSOR (EC 1.3.1.2) (DPD)  
DE (DIHYDROURACIL DEHYDROGENASE) (DIHYDROTHYMINE DEHYDROGENASE).  
GN DPYD.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Liver;  
RX MEDLINE=94365020; PubMed=8083224;  
RA Yokota H., Fernandez-Salguero P., Furuya H., Lin K., McBride O.W.,  
RA Podschun B., Schnackerz K.D., Gonzalez F.J.;  
RT "cDNA cloning and chromosome mapping of human dihydropyrimidine  
RT dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and  
RT congenital thymine uraciluria.";

RL J. Biol. Chem. 269:23192-23196(1994).  
RN [2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97280676; PubMed=9135003;  
RA Johnson M.R., Wang K., Tillmanns S., Albin N., Diasio R.B.;  
RT "Structural organization of the human dihydropyrimidine dehydrogenase  
RT gene.";  
RL Cancer Res. 57:1660-1663(1997).  
RN [3]  
RN SEQUENCE OF 581-635 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97047101; PubMed=8892022;  
RA Vreken P., van Kuilenburg A.B.P., Meinsma R., Smit G.P.A.,  
RA Bakker H.D., de Abreu R.A., van Gennip A.H.;  
RT "A point mutation in an invariant splice donor site leads to exon  
RT skipping in two unrelated Dutch patients with dihydropyrimidine  
RT dehydrogenase deficiency.";  
RL J. Inher. Metab. Dis. 19:645-654(1996).  
RN [4]  
RN CHARACTERIZATION, AND PARTIAL SEQUENCE.  
RC TISSUE=Liver;  
RX MEDLINE=92381021; PubMed=1512248;  
RA Lu Z.-H., Zhang R., Diasio R.B.;  
RT "Purification and characterization of dihydropyrimidine dehydrogenase  
RT from human liver.";  
RL J. Biol. Chem. 267:17102-17109(1992).  
RN [5]  
RN VARIANTS ARG-29; TRP-235 AND HIS-886.  
RX MEDLINE=96102836; PubMed=9439663;  
RA Vreken P., van Kuilenburg A.B.P., Meinsma R., van Gennip A.H.;  
RT "Dihydropyrimidine dehydrogenase (DPD) deficiency: Identification and  
RT expression of missense mutations C29R, R886H and R235W.";  
RL Hum. Genet. 101:333-338(1997).  
RN [6]  
RN VARIANTS ARG-29; TRP-235 AND HIS-886.  
RX MEDLINE=97411371; PubMed=9266349;  
RA Vreken P., van Kuilenburg A.B.P., Meinsma R., van Gennip A.H.;  
RT "Identification of novel point mutations in the dihydropyrimidine  
RT dehydrogenase gene.";  
RL J. Inher. Metab. Dis. 20:335-338(1997).  
CC -1- FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE  
CC REDUCTION OF URACIL AND THYMINE. ALSO INVOLVED THE DEGRADATION OF  
CC THE CHEMOTHERAPEUTIC DRUG 5-FLUOROURACIL.  
CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + NADP(+) -> URACIL + NADPH.  
CC -1- COFACTOR: TWO EACH OF FAD AND FMN. ALSO CONTAINS TWO 4FE-4S  
CC CLUSTERS. CONTAINS APPROXIMATELY 33 IRON ATOMS PER MOLECULE.  
CC -1- PATHWAY: INITIAL AND RATE-LIMITING ENZYME IN THE 3-STEP PATHWAY OF  
CC URACIL AND THYMINE CATABOLISM AND IN THE PATHWAY LEADING TO THE  
CC FORMATION OF BETA-ALANINE.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES WITH GREATEST ACTIVITY  
CC FOUND IN LIVER AND PERIPHERAL BLOOD MONONUCLEAR CELLS.  
CC -1- DISEASE: DEFECTS IN DPYD ARE THE CAUSE OF HEREDITARY THYMINE-  
CC URACILURIA (ALSO KNOWN AS FAMILIAL PYRIMIDINEMIA), A DISEASE  
CC CHARACTERIZED BY PERSISTENT URINARY EXCRETION OF EXCESSIVE AMOUNTS  
CC OF URACIL, THYMINE AND 5-HYDROXYMETHYLURACIL. PATIENTS SUFFERING  
CC FROM THIS DISEASE SHOW A SEVERE REACTION TO THE ANTICANCER DRUG 5-  
CC FLUOROURACIL. THIS REACTION INCLUDES STOMATITIS, LEUKOPENIA,  
CC THROMBOCYTOPENIA, HAIR LOSS, DIARRHEA, FEVER, MARKED WEIGHT LOSS,  
CC CEREBELLAR ATAXIA, AND NEUROLOGIC SYMPTOMS, PROGRESSING TO  
CC SEMICOMA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U09178; AAA57474.1; -  
DR EMBL; U20938; AAB51366.1; -

```

DR EMBL; X95670: CAA64973.1; ..
DR HSP; P26886; 1ERD.
DR MIM; 274270; ..
DR InterPro: IPR001295; ..
DR InterPro: IPR001450; ..
DR Pfam; PF01180; DHodehase; 1.
DR Pfam; PF00037; fer4; 3.
DR PROSITE; PS00198; 4FE4S-FERREDOXIN; 1.
KW Oxidoreductase; NADP; Flavoprotein; FAD; FMN; Iron-sulfur; 4Fe-4S;
FT PROPEP 1 3 DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+].
FT CHAIN 4 1025 NADP (POTENTIAL).
FT NP_BIND 335 331 FAD (POTENTIAL).
FT NP_BIND 471 481 FAD (POTENTIAL).
FT DOMAIN 661 678 URACIL-BINDING (POTENTIAL).
FT METAL 953 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 963 963 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 986 986 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT VARIANT 29 29 C->R (IN DPYD*9A AND DPYD*9B; LOSS OF ACTIVITY).
FT VARIANT 235 235 /FTIG-VAR_0051173.
FT VARIANT 534 534 R->W (IN DPYD*8; LOSS OF ACTIVITY).
FT VARIANT 534 534 /FTIG-VAR_0051174.
FT VARIANT 543 543 S->N (IN DPYD*4; LOW ACTIVITY).
FT VARIANT 543 543 /FTIG-VAR_0051175.
FT VARIANT 886 886 I->V (IN DPYD*5).
FT VARIANT 886 886 /FTIG-VAR_0051176.
FT VARIANT 995 995 R->H (IN DPYD*9B; 25% OF ACTIVITY).
FT VARIANT 910 910 V->F (IN DPYD*10; LOW ACTIVITY).
FT CONFLICT 910 910 /FTIG-VAR_0051178.
FT SEQUENCE 1025 AA; 111374 MW; 521C9430C7F69AFA CRC64;

Query Match 52.0%; Score 52.5; DB 1; Length 1025;
Best Local Similarity 63.2%; Pred. No. 0.7;
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPRKMKGLFSQAKISLF 20
DB 176 LPPPKMSEAYS-AKALF 193
||||| :|:|:|:|

RESULT 12
DPYD_BOVIN STANDARD; PRT; 1025 AA.
AC Q28007;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-NOV-2000 (Rel. 39, Last annotation update)
DE DIHYDROPYRIMIDINE DEHYDROGENASE [NADP+] (EC 1.3.1.2) (DPD)
DE (DIHYDROURACIL DEHYDROGENASE) (DIHYDROTHYMINE DEHYDROGENASE).
GN DPYD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97069948; PubMed=8912928;
RA Albin N., Johnson M.R., Diasio R.B.;
RT "cDNA cloning of bovine liver dihydropyrimidine dehydrogenase.";
RL DNA Seq. 6:243-250(1996).
CC -!- FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE
REDUCTION OF URACIL AND THYMINE.

```

```

CC -!- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + NADP(+) -> URACIL + NADPH.
CC -!- COFACTOR: FAD AND FMN. ALSO CONTAINS TWO 4FE-4S CLUSTERS.
CC -!- PATHWAY: INITIAL AND RATE-LIMITING ENZYME IN THE 3-STEP PATHWAY OF
URACIL AND THYMINE CATABOLISM AND IN THE PATHWAY LEADING TO THE
FORMATION OF BETA-ALANINE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U20981; AAB40985.1; ..
DR HSP; P26886; 1ERD.
DR InterPro: IPR001295; ..
DR InterPro: IPR001450; ..
DR Pfam; PF01180; DHodehase; 1.
DR Pfam; PF00037; fer4; 3.
DR PROSITE; PS00198; 4FE4S-FERREDOXIN; 1.
KW Oxidoreductase; NADP; Flavoprotein; FAD; FMN; Iron-sulfur; 4Fe-4S.
FT METAL 953 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 963 963 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
FT METAL 986 986 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
FT SEQUENCE 1025 AA; 111696 MW; 5B55F93A06C47F4F CRC64;

Query Match 45.0%; Score 45.5; DB 1; Length 1025;
Best Local Similarity 61.1%; Pred. No. 10;
Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPRKMKGLFSQAKISL 19
DB 176 LPPPKMPEAYS-AKAL 192
||||| :|:|:|:|

RESULT 13
RIR2_HSV11 STANDARD; PRT; 340 AA.
AC P10224;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE) (38 KDA SUBUNIT).
GN UL40.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574(1988).
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'DEoxyRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
THIOREDOXIN + H(2)O -> RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS.
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE

```

```

CC -----
CC SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10879; BAA01686.1; -.
CC EMBL; X14112; CAA32303.1; -.
CC PIR; D30088; WMBES7.
CC InterPro: IPR000358; -.
CC Pfam; PF00268; ribonuc_red; 1.
CC PROSITE; PS00368; RIBORED_SMALL; 1.
CC KW Oxidoreductase; DNA replication; Iron.
CC FT METAL 94 94 IRON 1 (BY SIMILARITY).
CC FT METAL 124 124 IRON 1 AND 2 (BY SIMILARITY).
CC FT METAL 127 127 IRON 1 (BY SIMILARITY).
CC FT METAL 187 187 IRON 2 (BY SIMILARITY).
CC FT METAL 221 221 IRON 2 (BY SIMILARITY).
CC FT METAL 224 224 IRON 2 (BY SIMILARITY).
CC FT ACT_SITE 131 131 BY SIMILARITY.
CC SQ SEQUENCE 340 AA; 38019 MW; 4B4ED94BF74FD3F CRC64;

Query Match 43.6%; Score 44; DB 1; Length 340;
Best Local Similarity 61.5%; Pred. No. 5.6;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQA 15
DB 242 PPDVRVGLFRQA 254

RESULT 14
RID_RSV1K
ID R1R2_HSV1K STANDARD; PRT; 340 AA.
AC P06474;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE) (38 KDA SUBUNIT).
GN UL40.
OS Herpes simplex virus (type 1 / strain KOS).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83059830; PubMed=6292456;
RA Draper K.G., Frink R.J., Wagner E.K.;
RT "Detailed characterization of an apparently unspliced beta herpes
RT simplex virus type 1 gene mapping in the interior of another.";
RL J. Virol. 43:1123-1128(1982).
RN [2]
RP REVISIONS.
RA Wagner E.K.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS.
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J02212; AAA66436.1; -.
CC InterPro: IPR000358; -.
CC Pfam; PF00268; ribonuc_red; 1.
CC PROSITE; PS00368; RIBORED_SMALL; 1.
CC KW Oxidoreductase; DNA replication; Iron.
CC FT METAL 94 94 IRON 1 (BY SIMILARITY).
CC FT METAL 124 124 IRON 1 AND 2 (BY SIMILARITY).
CC FT METAL 127 127 IRON 1 (BY SIMILARITY).
CC FT ACT_SITE 131 131 BY SIMILARITY.
CC SQ SEQUENCE 340 AA; 37966 MW; 921DC04B9D278DE5 CRC64;

Query Match 43.6%; Score 44; DB 1; Length 340;
Best Local Similarity 61.5%; Pred. No. 5.6;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQA 15
DB 242 PPDVRVGLFRQA 254

RESULT 15
DNAJ_METMA
ID DNAJ_METMA STANDARD; PRT; 389 AA.
AC P3515;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN DNAJ.
GN DNAJ.
OS Methanosarcina mazei.
OC Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93324351; PubMed=8332479;
RA Macario A.J.L., Dugan C.B., Clatens M., Conway de Macario E.;
RT "dnaJ in Archaea.";
RL Nucleic Acids Res. 21:2773-2773(1993).
CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60265; CAA42813.1; -.
CC PIR; S41748; S41748.
CC HSP; P25685; 1HDJ.
CC InterPro: IPR001305; -.
CC InterPro: IPR001623; -.
CC InterPro: IPR002939; -.
CC InterPro: IPR003095; -.
CC Pfam; PF00226; DnaJ; 1.
CC Pfam; PF01556; DnaJ_C; 1.
CC Pfam; PF00684; DnaJ_CXXCXGKG; 1.

```



DR PRINTS; PRO0625; DNAJPROTEIN.  
DR PROSITE; PS00636; DNAJ\_1; 1.  
DR PROSITE; PS0076; DNAJ\_2; 1.  
DR PROSITE; PS00637; DNAJ\_CXXCXGXG; 1.  
KW Chapterone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.  
FT DOMAIN 4 72 J-DOMAIN.  
FT REPEAT 73 110 GUY-RICH.  
FT REPEAT 144 151 CXXCXGXG MOTIF.  
FT REPEAT 161 168 CXXCXGXG MOTIF.  
FT REPEAT 187 194 CXXCXGXG MOTIF.  
FT REPEAT 201 208 CXXCXGXG MOTIF.  
FT METAL 144 144 ZINC 1 (BY SIMILARITY).  
FT METAL 147 147 ZINC 1 (BY SIMILARITY).  
FT METAL 161 161 ZINC 2 (BY SIMILARITY).  
FT METAL 164 164 ZINC 2 (BY SIMILARITY).  
FT METAL 187 187 ZINC 2 (BY SIMILARITY).  
FT METAL 190 190 ZINC 2 (BY SIMILARITY).  
FT METAL 201 201 ZINC 1 (BY SIMILARITY).  
FT METAL 204 204 ZINC 1 (BY SIMILARITY).  
SQ SEQUENCE 389 AA; 42990 MW; 5BE40E379D8C8139 CRC64;

Query Match 42.6%; Score 43; DB 1; Length 389;  
Best Local Similarity 66.7%; Pred. No. 9.4;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 5 PRKMKGLFSQAK 16  
Db 373 PRKSKGLFEKVK 384  
||| |||| :|

Search completed: April 24, 2001, 16:34:00  
Job time: 556 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:33:08 ; Search time 113.97 Seconds  
(without alignments)  
20.568 Million cell updates/sec

Title: US-09-171-432A-42

Perfect score: 101  
Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_15.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_unclassified.\*  
13: sp\_vertebrate.\*  
14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	101	100.0	116	14	071977	071977	hepatitis a
2	101	100.0	116	14	Q9W7X7	Q9W7X7	hepatitis a
3	101	100.0	116	14	Q9W7S7	Q9W7S7	hepatitis a
4	101	100.0	116	14	Q92941	Q92941	hepatitis a
5	101	100.0	1124	14	Q84780	Q84780	hepatitis a
6	101	100.0	1161	14	Q05794	Q05794	hepatitis a
7	101	100.0	2227	14	Q9WMA1	Q9WMA1	hepatitis a
8	101	100.0	2227	14	Q9WMA9	Q9WMA9	hepatitis a
9	101	100.0	2227	14	Q9IFH5	Q9IFH5	hepatitis a
10	99	98.0	116	14	Q9W7X8	Q9W7X8	hepatitis a
11	98	97.0	112	14	Q86534	Q86534	hepatitis a
12	98	97.0	116	14	Q71975	Q71975	hepatitis a
13	98	97.0	116	14	Q71978	Q71978	hepatitis a
14	98	97.0	116	14	Q9W7T4	Q9W7T4	hepatitis a
15	98	97.0	2216	14	Q9WMA2	Q9WMA2	hepatitis a
16	98	97.0	2218	14	Q67824	Q67824	hepatitis a
17	98	97.0	2218	14	Q67817	Q67817	hepatitis a
18	98	97.0	2218	14	Q67825	Q67825	hepatitis a
19	98	97.0	2227	14	Q67826	Q67826	hepatitis a

20	98	97.0	2227	14	Q9WMA3	Q9WMA3	hepatitis a
21	97	96.0	2227	14	Q9WMA4	Q9WMA4	hepatitis a
22	95	94.1	116	14	Q9W7T0	Q9W7T0	hepatitis a
23	95	94.1	2227	14	Q9WMA0	Q9WMA0	hepatitis a
24	94	93.1	115	14	Q71976	Q71976	hepatitis a
25	94	93.1	184	14	Q87092	Q87092	simian hepa
26	48	47.5	434	10	Q9LH93	Q9LH93	arabidopsis
27	48	47.5	434	10	Q9LH92	Q9LH92	arabidopsis
28	47	46.5	414	10	Q9M8M9	Q9M8M9	arabidopsis
29	46.5	46.0	1025	11	Q89000	Q89000	rattus norv
30	45	44.6	178	10	Q9LUD4	Q9LUD4	arabidopsis
31	45	44.6	519	5	Q9VUI9	Q9VUI9	drosofila
32	44	43.6	113	10	Q9ZSE1	Q9ZSE1	pinus radia
33	44	43.6	343	2	P72959	P72959	synecocyst
34	44	43.6	345	10	Q9M3U4	Q9M3U4	vitis vinif
35	44	43.6	355	10	Q9LW46	Q9LW46	arabidopsis
36	44	43.6	369	2	P73589	P73589	synecocyst
37	44	43.6	775	10	Q9LXV8	Q9LXV8	arabidopsis
38	43	42.6	309	1	Q9UYA6	Q9UYA6	pyrococcus
39	43	42.6	379	10	Q9LYX2	Q9LYX2	arabidopsis
40	43	42.6	587	2	Q53609	Q53609	streptomyce
41	42.5	42.1	105	1	Q9YCK7	Q9YCK7	aeropyrum p
42	42	41.6	109	3	O13543	O13543	saccharomyc
43	42	41.6	117	5	O61138	O61138	cryptospori
44	42	41.6	165	9	Q38430	Q38430	bacterioph
45	42	41.6	178	10	O64699	O64699	arabidopsis

ALIGNMENTS

RESULT 1  
ID 071977 PRELIMINARY; PRT; 116 AA.  
AC 071977;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-16; TISSUE-SERUM;  
RX MEDLINE-97442186; PubMed-9298735;  
RA Grinde B., Stene-Johansen K., Sharma B., Hoel T., Jensenius M.,  
RA Skaug K.;  
RT "Characterisation of an epidemic of hepatitis A virus involving  
RT intravenous drug abusers--infection by needle sharing?";  
RL J. Med. Virol. 53:69-75(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-16; TISSUE-SERUM;  
RA Stene-Johansen K., Skaug K., Blystad H., Grinde B.;  
RL Scand. J. Infect. Dis. 0:0-0(1998).  
DR EMBL; AF050237; AAC05530.1;  
KW Polyprotein.  
FT NON\_TER 1 116  
FT NON\_TER 116 116  
SQ SEQUENCE 116 AA; 13493 MW; 8E26DEB9E7EFB77 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 116;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 20  
|||||  
Db 96 VLPPPRKMKGLFSQAKISLF 115



```

RA Blinov V.M., Vasilenko S.K., Sandakhchiev L.S., Kusov Y.Y.,
RL Gradko V.I., Flier G.P., Balyan M.S., Drozdov S.G.;
RL Dokl. Biochem. 285:379-383(1986).
DR EMBL; X04200; CAA27797.1; -.
DR EMBL; A11312; CAA00953.1; -.
KW Nonstructural protein.
FT NON_TER 1124 1124
FT NON_TER 1124 1124
SQ SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

Query Match 100.0%; Score 101; DB 14; Length 1124;
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQAKISLF 20
   |||||
DB 780 VLPPrKMKGLFSQAKISLF 799

RESULT 6
Q05794 PRELIMINARY; PRT; 1161 AA
AC Q05794; Q67800; Q67801; Q67802; Q67803; Q67804; Q67805; Q67806;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D.(EC 2.7.7.48)) (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilenko S.K.,
RA Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,
RA Balayan M.S.;
RL Mol. Gen. Microbiol. Virol. 6:129-133(1987).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL; X15464; CAA33492.1; -.
DR INTERPRO; IPR000408; -.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein; Coat protein; Core protein; RNA-directed RNA polymerase;
KW Hydrolase; Thiol protease.
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).
FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).
FT CHAIN 795 900 CORE PROTEIN P2A.
FT CHAIN 901 1087 CORE PROTEIN P2B.
FT CHAIN 1088 >1161 CORE PROTEIN P2C.
FT NON_TER 1161 1161
SQ SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 1161;
Best Local Similarity 100.0%; Pred. No. 2.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQAKISLF 20
   |||||
DB 817 VLPPrKMKGLFSQAKISLF 836

RESULT 7
Q9WNA1 PRELIMINARY; PRT; 2227 AA
ID Q9WNA1
AC Q9WNA1
DT 01-NOV-1999 (Tremblrel. 12, Created)

```

```

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FH3;
RA Fujiwara K.;
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB020567; BAA35105.1; -.
DR INTERPRO; IPR000408; -.
DR INTERPRO; IPR000605; -.
DR INTERPRO; IPR001205; -.
DR INTERPRO; IPR001643; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSN.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251415 MW; F92C8E2323FC5621 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQAKISLF 20
   |||||
DB 823 VLPPrKMKGLFSQAKISLF 842

RESULT 8
Q9WM99 PRELIMINARY; PRT; 2227 AA
ID Q9WM99
AC Q9WM99
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FH3;
RA Fujiwara K.;
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB020569; BAA35107.1; -.
DR INTERPRO; IPR000408; -.
DR INTERPRO; IPR000605; -.
DR INTERPRO; IPR001205; -.
DR INTERPRO; IPR001643; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSN.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251387 MW; 81764018672A23D CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPrKMKGLFSQAKISLF 20
   |||||
DB 823 VLPPrKMKGLFSQAKISLF 842

```

```
RESULT 9
Q91FH5 Q91FH5 PRELIMINARY; PRT; 2227 AA.
AC Q91FH5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYPROTEIN PRECURSOR.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HAF-203;
RA Baptista M.L., Silva M., de Lima M.A., Yoshida C.F., Gaspar A.M.,
RA Pires Lopes M.Q., Galler R.;
RT "Nucleotide sequence of the HAF-203 hepatitis A virus strain isolated
RT in Brazil and expression of the VP1 gene in a bacterial system.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF268396; AAF80114.1; -.
SQ SEQUENCE 2227 AA; 251432 MW; 81913AEC8A04200 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20
Db 823 VLPPrKMKGLFSQAKISLF 842
|||||
RESULT 10
Q9W7X8 Q9W7X8 PRELIMINARY; PRT; 116 AA.
AC Q9W7X8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOR-13; TISSUE=SERUM;
RX MEDLINE=97442186; PubMed=9298735;
RA Grinde B., Stene-Johansen K., Sharma B., Hoel T., Jensenius M.,
RA Skaug K.;
RT "Characterisation of an epidemic of hepatitis A virus involving
RT intravenous drug abusers--infection by needle sharing?";
RL J. Med. Virol. 53:69-75(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NOR-13; TISSUE=SERUM;
RA Stene-Johansen K., Skaug K., Blystad H., Grinde B.;
RT "An unique hepatitis A virus strain caused an epidemic in Norway
RT associated with intravenous drug abuse.";
RL Scand. J. Infect. Dis. 0:0-0(1998).
DR EMBL; AF050234; AAC05527.1; -.
KW Polyprotein.
FT NON_TER 1 116
SQ SEQUENCE 116 AA; 13507 MW; 8E26DBE9E7EBA77 CRC64;

Query Match 98.0%; Score 99; DB 14; Length 116;
Best Local Similarity 95.0%; Pred. No. 5.5e-09;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VLPPrKMKGLFSQAKISLF 20
Db 96 VLPPrKMKGLFSQAKISLF 115
|||||
RESULT 11
Q86534 Q86534 PRELIMINARY; PRT; 112 AA.
AC Q86534;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE VP1 PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92348853; PubMed=1668326;
RA Fineschi N., Cavallieri F., Garelick H., Prugnola A., Pellegrini V.,
RA Zuckerman A.J.;
RT "Characterization of a hepatitis A virus strain suitable for vaccine
RT production.";
RL J. Hepatol. 13 Suppl. 4:146-151(1991).
DR EMBL; S44105; CAB28369.1; -.
FT NON_TER 112 112
SQ SEQUENCE 112 AA; 13145 MW; 3C0CB8B4570D9A2C CRC64;

Query Match 97.0%; Score 98; DB 14; Length 112;
Best Local Similarity 95.0%; Pred. No. 7.8e-09;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20
Db 64 VLPPrKMKGLFSQAKISLF 83
|||||
RESULT 12
Q71975 Q71975 PRELIMINARY; PRT; 116 AA.
AC Q71975;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOR-5; TISSUE=SERUM;
RX MEDLINE=97442186; PubMed=9298735;
RA Grinde B., Stene-Johansen K., Sharma B., Hoel T., Jensenius M.,
RA Skaug K.;
RT "Characterisation of an epidemic of hepatitis A virus involving
RT intravenous drug abusers--infection by needle sharing?";
RL J. Med. Virol. 53:69-75(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NOR-5; TISSUE=SERUM;
RA Stene-Johansen K., Skaug K., Blystad H., Grinde B.;
RL Scand. J. Infect. Dis. 0:0-0(1998).
DR EMBL; AF050227; AAC05520.1; -.
KW Polyprotein.
FT NON_TER 1 116
SQ SEQUENCE 116 AA; 13461 MW; 6E37D34AFA440D6A CRC64;
```

Query Match 97.0%; Score 98; DB 14; Length 116;  
Best Local Similarity 95.0%; Pred. No. 8.1e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 96 VLPPPRKMKGVFSQAKISLF 115

RESULT 13  
ID 071978 PRELIMINARY; PRT; 116 AA.  
AC 071978;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-17; TISSUE-SERUM;  
RX MEDLINE-97442186; PubMed=9298735;  
RA Grinde B., Stene-Johansen K., Sharma B., Hoel T., Jensenius M.,  
RA Skaug K.;  
RT "Characterisation of an epidemic of hepatitis A virus involving  
RT intravenous drug abusers--infection by needle sharing?";  
RL J. Med. Virol. 53:69-75(1997).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-17; TISSUE-SERUM;  
RA Stene-Johansen K., Skaug K., Blystad H., Grinde B.;  
RL Scand. J. Infect. Dis. 0:0-0(1998).  
DR EMBL; AF050238; AAC05531.1;  
KW Polyprotein.  
FT NON\_TER 1 1  
FT NON\_TER 116 116  
SQ SEQUENCE 116 AA; 13479 MW; 8E26C24BE7FBB77 CRC64;

Query Match 97.0%; Score 98; DB 14; Length 116;  
Best Local Similarity 95.0%; Pred. No. 8.1e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 96 VLPPPRKMKGVFSQAKISLF 115

RESULT 14  
ID 09W7T4 PRELIMINARY; PRT; 116 AA.  
AC 09W7T4;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-14, NOR-2; TISSUE-SERUM;  
RX MEDLINE-97442186; PubMed=9298735;  
RA Grinde B., Stene-Johansen K., Sharma B., Hoel T., Jensenius M.,  
RA Skaug K.;  
RT "Characterisation of an epidemic of hepatitis A virus involving  
RT intravenous drug abusers--infection by needle sharing?";  
RL J. Med. Virol. 53:69-75(1997).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-NOR-14, NOR-2; TISSUE-SERUM;  
RA Stene-Johansen K., Skaug K., Blystad H., Grinde B.;  
RT "An unique hepatitis A virus strain caused an epidemic in Norway  
RT associated with intravenous drug abuse";  
RL Scand. J. Infect. Dis. 0:0-0(1998).  
DR EMBL; AF050235; AAC05528.1;  
DR EMBL; AF050224; AAC05517.1;  
KW Polyprotein.  
FT NON\_TER 1 1  
FT NON\_TER 116 116  
SQ SEQUENCE 116 AA; 13493 MW; 8E26C34AE7EBA77 CRC64;

Query Match 97.0%; Score 98; DB 14; Length 116;  
Best Local Similarity 95.0%; Pred. No. 8.1e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 96 VLPPPRKMKGVFSQAKISLF 115

RESULT 15  
ID 09WMA2 PRELIMINARY; PRT; 2216 AA.  
AC 09WMA2;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AH3;  
RA Fujiwara K.;  
RT "Hepatitis A virus";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020566; BAA35104.1;  
DR INTERPRO; IPR000408;  
DR INTERPRO; IPR000605;  
DR INTERPRO; IPR001205;  
DR INTERPRO; IPR001643;  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUSNS.  
DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2216 AA; 250209 MW; 1A9D93FEC21FBE82 CRC64;

Query Match 97.0%; Score 98; DB 14; Length 2216;  
Best Local Similarity 95.0%; Pred. No. 1.6e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 823 VLPPPRKMKGVFSQAKISLF 842

Search completed: April 24, 2001, 16:33:10  
Job time: 571 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:23:22 ; Search time 121.9 seconds  
(without alignments)  
9.379 Million cell updates/sec

Title: US-09-171-432a-43

Perfect score: 101

Sequence: 1 FSQAKISLFTEHEIMKFS 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	20	18 W42926	Immunogenic Hepati
2	101	100.0	854	6 P50287	Sequence encoded b
3	101	100.0	1077	20 W95559	A partial hepatiti
4	101	100.0	1091	14 R32426	Translated from 5'
5	101	100.0	2227	11 R05697	Attenuated hepatit
6	101	100.0	2227	18 W34074	Hepatitis A virus
7	101	100.0	2227	21 B18607	Amino acid sequenc
8	101	100.0	2227	21 B18608	Amino acid sequenc
9	101	100.0	2227	21 B18609	Amino acid sequenc
10	98	97.0	2227	7 P60066	Sequence of viral
11	97	96.0	366	6 P50230	Sequence of hepati

12	97	96.0	993	6	P50116	Sequence of Hepati
13	97	96.0	993	6	P50231	Sequence encoded b
14	49.5	49.0	20	19	W76539	Plasmid pGEM2-HAV-
15	47	46.5	518	20	Y27340	Group B Streptococ
16	47	46.5	816	20	Y27339	Group B Streptococ
17	42	41.6	20	18	W42925	Immunogenic Hepati
18	41	40.6	173	22	B49279	Protein encoded by
19	41	40.6	2431	13	R25138	SFV4 non-structura
20	40	39.6	105	21	G19719	Arabidopsis thalia
21	40	39.6	105	21	G61014	Arabidopsis thalia
22	40	39.6	122	21	G41997	Arabidopsis thalia
23	40	39.6	173	22	B49278	Protein encoded by
24	39	38.6	148	20	X35073	Chlamydia pneumoni
25	39	38.6	170	20	X33564	G. max CBF homolog
26	39	38.6	289	13	R30190	UGT1E Exon 1 produ
27	39	38.6	289	21	Y57095	UDP-glucuronosyltr
28	39	38.6	455	20	X34954	Chlamydia pneumoni
29	39	38.6	750	15	R47858	Human LDL receptor
30	39	38.6	800	11	R07713	Human low density
31	39	38.6	839	22	B49601	Sequence of human
32	39	38.6	860	15	R47157	Human LDL receptor
33	39	38.6	860	15	R47860	Chicken p95/human
34	39	38.6	924	16	R78234	LDLR/TF chimeric p
35	39	38.6	1410	18	W07622	Delta-endotoxin 20
36	38	37.6	265	16	R70734	Delta-endotoxin.
37	38	37.6	265	17	R91966	B. thuringiensis p
38	38	37.6	265	18	W07850	CytC toxin CYTC876
39	38	37.6	265	20	Y01207	rii-2 protein. Es
40	38	37.6	265	21	R03646	A. pleuropneumonia
41	38	37.6	319	14	R32193	Arabidopsis thalia
42	38	37.6	707	18	W22153	Arabidopsis thalia
43	38	37.6	707	21	Y51407	Arabidopsis thalia
44	37.5	37.1	309	21	G15792	Arabidopsis thalia
45	37.5	37.1	309	21	G42809	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1  
W42926  
ID W42926 standard; peptide; 20 AA.  
XX  
AC W42926;  
XX  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1319.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW Immune response; antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
PN W09740147-Al.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US06891.  
XX  
PR 19-APR-1996; 96US-0015644.  
XX  
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an  
PT Immune response to HAV in a mammal or to detect the presence of  
PT antibodies against HAV in a mammal  
XX

PS Claim 18; Page 112; 140pp; English.

XX Peptides W42922-30 are immunogenic peptides corresponding to immunogenic  
 CC epitopes of the Hepatitis A virus (HAV). The peptides are substantially  
 CC similar to a portion of the amino acid sequence of the p2A protein of HAV  
 CC corresponding to amino acids 792-980. The present peptide is derived  
 CC from amino acids 834-853, and has a reactivity of 27.1% with acute sera.  
 CC Compositions containing the peptides can be used to induce an immune  
 CC response to HAV in a mammal. The peptides can also be used to detect the  
 CC presence of antibodies against HAV in mammalian serum. The peptides can  
 CC also be used to make an antibody against HAV by administering the peptide  
 CC to a mammal.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 101; DB 18; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20

Db 1 fsqakislftyteeheimkfs 20

RESULT 2

P50287  
 ID P50287 standard; Protein; 854 AA.

XX AC P50287;

XX 30-NOV-1991 (first entry)

DE Sequence encoded by hepatitis A virus (HAV) cDNA from near the  
 DE genome 5' terminus to the end of the area corresponding to the  
 DE capsid protein region of poliovirus RNA.

XX Hepatitis A virus assay; antigen; antibody.

OS Hepatitis A virus.

PN W08501517-A.

XX 11-APR-1985.

XX 27-SEP-1984; 84WO-US01552.

XX 30-SEP-1983; 83US-0537911.

XX (MASI ) MASSACHUSETTS INST TECH.

XX Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;

PI Racanietello VR;

XX WPI; 1985-098846/16.

DR N-PSDB; N50330.

XX New hepatitis A virus cDNA - useful in assays for the virus and  
 PT for prodn. of the viral antigen and antibodies to it

XX Example; Fig 7; 60pp; English.

XX The inventors claim HAV cDNA and a method for producing it, whereby  
 CC large amts. can be obtd. economically. The cDNA is useful in the  
 CC assay for detection of HAV quickly and easily and with high  
 CC sensitivity and specificity. The HAV cDNA is also used in the prodn.  
 CC of HAV antigen or antibodies to it. The antibodies may be monoclonal.

XX SQ Sequence 854 AA;

Query Match 100.0%; Score 101; DB 6; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20

Db 834 fsqakislftyteeheimkfs 853

RESULT 3

W95559

ID W95559 standard; Protein; 1077 AA.

XX AC W95559;

XX 28-APR-1999 (first entry)

DE A partial hepatitis A virus (HAV) protein.

XX Hepatitis A virus protein; HAV; P2 region;  
 KW cell-culture-adapted HAV strain; infection; accelerated growth.

OS Hepatitis A virus.

XX US5849562-A.

XX 15-DEC-1998.

XX 06-JUN-1995; 95US-0468926.

XX 06-NOV-1991; 91US-0788262.

XX 30-SEP-1983; 83US-0537911.

XX 27-SEP-1984; 84US-0654942.

XX 06-OCT-1988; 88US-0256135.

XX 06-JUN-1995; 95US-0468926.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Emerson SU, Purcell RH;

XX WPI; 1999-094412/08.

XX N-PSDB; X01006.

XX Chimeric hepatitis A virus strains - with P2 region from  
 PT cell-culture-adapted strain in wild-type genome

XX Disclosure; Fig 7A-L; 36pp; English.

XX The present sequence represents a partial hepatitis A virus (HAV)

CC protein. The specification describes a DNA construct consisting

CC of a wild-type HAV genome in which the P2 region is replaced by the

CC P2 region from a cell-culture-adapted HAV strain. The construct is

CC used to demonstrate that mutations in the P2 region of a

CC cell-culture-adapted HAV strain are sufficient for establishment of

CC infection and accelerated growth in cell culture.

XX SQ Sequence 1077 AA;

Query Match 100.0%; Score 101; DB 20; Length 1077;

Best Local Similarity 100.0%; Pred. No. 2.3e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20

Db 1057 fsqakislftyteeheimkfs 1076

RESULT 4

R32426

ID R32426 standard; Protein; 1091 AA.

XX AC R32426;

XX 10-JUN-1993 (first entry)

XX Translated from 5' region of Hepatitis A Virus genomic clone.  
 DE HAV HM-175; chronic liver disease; picornavirus.  
 KW Hepatitis A Virus.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Region 238..1091  
 FT /label= ORF  
 FT /note= "second putative initiation codon at  
 FT position 240"  
 FT Region 1..711  
 FT /note= "X's correspond to nonsense codons,  
 FT i.e. this region is not an ORF"  
 PN US7788262-A.  
 XX  
 PD 15-DEC-1992.  
 XX  
 XX 30-SEP-1983; 83US-0536911.  
 XX  
 XX 27-SEP-1984; 84US-0654942.  
 PR 06-OCT-1988; 88US-0256135.  
 PR 30-SEP-1983; 83US-0536911.  
 PR 06-NOV-1991; 91US-0788262.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 XX  
 PI Baltimore D, Feinstein SM;  
 PI Purcell RH, Racaniello VR, Ticehurst JR;  
 XX  
 DR WPI; 1993-067429/08.  
 DR N-PSDB; Q36934.  
 XX  
 PT Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn.  
 PT of antigen and antibodies  
 XX  
 PS Disclosure; Fig 7; 65pp; English.  
 XX  
 CC HAV virion RNA was extracted from the livers of marmosets which had  
 CC been inoculated with HAV (the HAV had previously been passaged twice  
 CC in marmosets). The RNA was used to prepare ds cDNA clones by  
 CC standard methods. Clones contg. inserts which hybridised to RNA from  
 CC HAV-infected African Green Monkey Kidney cells were selected for  
 CC further analysis. A 7.4kb restriction map (about 99% of the HAV  
 CC genome) was constructed from 5 overlapping inserts. The sequence of  
 CC the first 3.3kb (approx.) from the 5'-terminus was determined. An  
 CC amino acid sequence was decoded from the entire clone and an open  
 CC reading frame was identified starting at position 238. A comparison  
 CC of the predicted HAV amino acid sequences with the known capsid  
 CC protein sequences of other picornaviruses (poliovirus, foot and  
 CC mouth disease virus and encephalomyelitis virus) revealed areas of  
 CC local homology.  
 XX  
 SQ Sequence 1091 AA;  
 Query Match 100.0%; Score 101; DB 14; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FSQAKISLFYTEEHMKFS 20  
 Db 1071 fsqakislfyteeheimkfs 1090  
 RESULT 5  
 ID R05697  
 AC R05697 standard; protein; 2227 AA.  
 XX R05697;  
 XX

DT 15-AUG-1990 (first entry)  
 XX  
 DE Attenuated hepatitis A virus.  
 XX  
 KW Hepatitis A virus; vaccine; attenuated.  
 XX  
 OS Hepatitis A virus, strain HM-175.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..23  
 FT /label=VP4 = 1A  
 FT Region 24..245  
 FT /label=VP2 = 1B  
 FT Region 246..491  
 FT /label=VP3 = 1C  
 FT Region 492..791  
 FT /label=VP1 = 1D  
 FT Region 792..980  
 FT /label=2A  
 FT Region 981..1087  
 FT /label=2B  
 FT Region 1088..1422  
 FT /label=2C  
 FT Region 1423..1496  
 FT /label=3A  
 FT Region 1497..1519  
 FT /label=3B = VPg  
 FT Region 1520..1738  
 FT /label=3C  
 FT Region 1739..2227  
 FT /label=3D  
 XX  
 PN US4894228-A.  
 XX  
 PD 16-JAN-1990.  
 XX  
 PF 12-JUL-1988; 88US-0217824.  
 XX  
 PR 12-JUL-1988; 88US-0217824.  
 PR 12-JUL-1988; 88US-0652967.  
 XX  
 XX (USSH ) US DEPT HEALTH & HUMAN.  
 PA Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;  
 PI Daemer RJ, Gust ID;  
 XX  
 DR WPI; 1990-075557/10.  
 DR N-PSDB; Q03512.  
 XX  
 PT Vaccine against hepatitis A virus infection - comprises novel  
 PT attenuated hepatitis A virus strain.  
 XX  
 PS Claim 1; Fig 1; 18pp; English.  
 XX  
 CC The attenuated HAV is useful for inducing protective immunity against  
 CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
 CC several nucleotide changes distributed throughout the genome, is  
 CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
 CC suitable for use as an HAV vaccine. It is noted that not all the changes  
 CC are necessary for attenuation and use as a vaccine.  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 101; DB 11; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FSQAKISLFYTEEHMKFS 20  
 Db 834 fsqakislfyteeheimkfs 853



```

B18608
ID B18608 standard; Protein: 2227 AA.
XX
XX AC B18608;
XX
XX DT 15-JAN-2001 (first entry)
XX
XX DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.
XX
XX KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
XX P-35 virus.
XX
XX OS Hepatitis A virus.
XX
XX PN US6113912-A.
XX
XX PD 05-SEP-2000.
XX
XX PF 07-JUN-1995; 95US-0475886.
XX
XX PR 18-SEP-1992; 92US-0947338.
XX 17-SEP-1993; 93WO-US08610.
XX 10-MAR-1995; 95US-0397232.
XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
XX DR WPI; 2000-586464/55.
XX N-PSDB; A75478.
XX
XX PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
XX line useful as vaccine for protecting humans against hepatitis A virus
XX infection, has modified genome compared to wild type
XX
XX PS Disclosure: Columns 93-104; 72pp; English.
XX
XX CC The present sequence is derived from a live attenuated hepatitis A
XX virus (HAV) of the invention, designated HAV 4380. The sequence is
XX produced by modifying wild type HAV strain HM-174. The HAV of the
XX invention are adapted to growth in the human fibroblast-like cell
XX line MRC-5. The HAV is able to propagate in MRC-5 cells and retain
XX appropriate attenuation. It is useful as a live vaccine for prophylaxis
XX of hepatitis A in humans and other primates.
XX
XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHIMKFS 20
Db 834 fsqakislfyteeheimkfs 853

RESULT 10
P60066
ID P60066 standard; Protein: 2227 AA.
XX
XX AC P60066;
XX
XX DT 26-JUN-1991 (first entry)
XX
XX DE Sequence of viral I434 polypeptide encoded by the complete
XX nucleotide sequence of the HAV genome.
XX
XX KW Diagnosis; vaccine; passive immunotherapy.
XX
XX OS Hepatitis A virus.
XX
XX FH Key Location/Qualifiers
XX FT Region 1..245
XX FT /label= P1.1A
XX FT Region 246..491
XX FT /label= 1B
XX FT Region 492..836
XX FT /label= 1C
XX FT Region 837..980
XX FT /label= P2.2A
XX FT Region 981..1076
XX FT /label= 2B
XX FT Region 1077..1422
XX FT /label= 2C
XX FT Region 1423..1484
XX FT /label= P3.3A
XX FT Region 1485..1507

```

---

```

B18609
ID B18609 standard; Protein: 2227 AA.
XX
XX AC B18609;
XX
XX DT 15-JAN-2001 (first entry)
XX
XX DE Amino acid sequence of live attenuated Hepatitis A virus 4380.
XX
XX KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
XX HAV 4380.
XX
XX OS Hepatitis A virus.
XX
XX PN US6113912-A.
XX
XX PD 05-SEP-2000.
XX
XX PF 07-JUN-1995; 95US-0475886.
XX
XX PR 18-SEP-1992; 92US-0947338.
XX 17-SEP-1993; 93WO-US08610.
XX 10-MAR-1995; 95US-0397232.
XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
XX DR WPI; 2000-586464/55.
XX N-PSDB; A75477.
XX
XX PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
XX line useful as vaccine for protecting humans against hepatitis A virus
XX infection, has modified genome compared to wild type
XX
XX PS Disclosure: Columns 67-78; 72pp; English.
XX
XX CC The present sequence is derived from passage 35 of a wild type
XX hepatitis A virus (HAV) strain HM-174. The resulting virus is
XX designated P-35 virus. The sequence is modified to produce HAV which
XX are adapted to growth in the human fibroblast-like cell line MRC-5.
XX The HAV is able to propagate in MRC-5 cells and retain appropriate
XX attenuation. It is useful as a live vaccine for prophylaxis of
XX hepatitis A in humans and other primates.
XX
XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHIMKFS 20
Db 834 fsqakislfyteeheimkfs 853

RESULT 9
B18609
ID B18609 standard; Protein: 2227 AA.
XX
XX AC B18609;
XX
XX DT 15-JAN-2001 (first entry)
XX
XX DE Amino acid sequence of live attenuated Hepatitis A virus 4380.
XX
XX KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
XX HAV 4380.
XX
XX OS Hepatitis A virus.
XX
XX PN US6113912-A.

```

FT Region /label= 3B  
 FT 1508..1678  
 FT /label= 3C  
 FT 1679..2227  
 FT /label= 3D  
 XX  
 PN EPI99480-A.  
 XX  
 PD 29-OCT-1986.  
 XX  
 XX  
 PF 03-APR-1986; 86EP-0302465.  
 XX  
 PR 03-APR-1985; 85US-0719329.  
 XX  
 PA (CHIR-) CHIRON CORP.  
 XX  
 PI Dina D, Potter SJ, Vannest GA, Caput D;  
 XX  
 DR WPI; 1986-286213/44.  
 DR N-PSDB; N60080.  
 XX  
 PT Hepatitis A virus nucleotide sequence and polypeptide - and use  
 PT in prodn. of vaccines and diagnostic probes  
 XX  
 PS Claim 5; Fig 1; 18pp; English.  
 XX  
 CC N60080 and oligonucleotide fragments are useful in detection of  
 CC hepatitis A virus; transformed hosts may be used for expression of  
 CC polypeptides and fragments useful in vaccines without risk of  
 CC infection by the virus or in prodn. of particles which are capable  
 CC of inducing immunocompetent B cells for passive immunotherapy. Pref.  
 CC epitope is derived from AAs 445-657 or 792-848 of the HAV  
 CC polypeptide sequence (P60066).  
 XX  
 SQ Sequence 2227 AA;

Query Match 97.0%; Score 98; DB 7; Length 2227;  
 Best Local Similarity 95.0%; Pred. No. 1.6e-07;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEHEIMKFS 20  
 |||||||||:|||||||  
 Db 834 fsqakislfydeheimkfs 853

RESULT 11  
 P50230  
 ID P50230 standard; Protein; 366 AA.  
 XX  
 AC P50230;  
 XX  
 DT 28-NOV-1991 (first entry)  
 XX  
 DE Sequence of hepatitis A virus (HAV) surface protein (VP-1).  
 XX  
 KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
 KW diagnostic assay.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN EPI38704-A.  
 XX  
 PD 24-APR-1985.  
 XX  
 PF 09-OCT-1984; 84EP-0402025.  
 XX  
 PR 02-MAR-1984; 84US-0585942.  
 PR 14-OCT-1983; 83US-0541836.  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 PI Hughes JV, Scolnick EM, Tomassini JE;

XX WPI; 1985-100818/17.  
 DR N-PSDB; N50274.  
 XX  
 PT New hepatitis A virus surface protein - useful for binding to  
 PT neutralising antibodies to the virus  
 XX  
 PS Claim 21; Page 46-48; 49pp; English.  
 XX  
 CC VPI is isolated by solubilisation of the intact virus in an aq.  
 CC anionic surfactant and a reducing agent. The viral proteins are sepd.  
 CC and the protein of molecular wt. 33000 daltons is sepd.  
 XX  
 SQ Sequence 366 AA;

Query Match 96.0%; Score 97; DB 6; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-08;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEHEIMKF 19  
 |||||||||:|||||||  
 Db 343 fsqakislfyteeheimkf 361

RESULT 12  
 P50116  
 ID P50116 standard; Protein; 993 AA.  
 XX  
 AC P50116;  
 XX  
 DT 30-SEP-1991 (first entry)  
 XX  
 DE Sequence of Hepatitis A virus (HAV) immunogenic peptides  
 DE VP-1, VP-2, VP-3 and VP-4.  
 XX  
 KW Antigenic protein; immunogen; vaccine.  
 XX  
 OS Hepatitis A virus (strain CR326).  
 XX  
 PN EPI54587-A.  
 XX  
 PD 11-SEP-1985.  
 XX  
 PF 27-FEB-1985; 85EP-0400369.  
 XX  
 PR 02-MAR-1984; 84US-0585818.  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 PI Linemeyer DL, Menke JG, Reuben RG, Mitra SW;  
 XX  
 DR WPI; 1985-224964/37.  
 DR N-PSDB; N50139.  
 XX  
 PT New nucleotide sequences coding for hepatitis A virus antigens -  
 PT useful for eliciting normal immune response and in vaccines for  
 PT protecting against the virus  
 XX  
 PS Example; Page 11-17; 32pp; English.  
 XX  
 CC Within the sequence in N50139 is encoded the information necessary  
 CC to make the antigenic proteins of HAV. The sequences encoding for  
 CC the structural proteins begin at base 403. The key sub-unit  
 CC sequences within VP-1, designated Sequences I,II,III,IV, and V,  
 CC start, respectively at 1882, 1963, 1999, 2146, 2347. Other  
 CC nucleotide sequences which are valuable as encoding antigenic  
 CC proteins are the sequences from base 1749 to base 2722; from base  
 CC 1487 to base 2980 and from base 1644 to base 2722. The sequence from  
 CC base 1749 to base 2722 is esp. valuable as a vector for producing  
 CC antigen protein. Sequences II-V are claimed. X in P50116 denotes the  
 CC translation of a stop codon.  
 XX

SQ Sequence 993 AA;

Query Match 96.0%; Score 97; DB 6; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-08;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHEIMKF 19  
 |||||  
 Db 970 fsqakislfyteeheimkf 988

RESULT 13

P50231  
 ID P50231 standard; Protein; 993 AA.

AC P50231;

XX 28-NOV-1991 (first entry)

DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
 including surface protein (VP-1).

KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
 diagnostic assay.

OS Hepatitis A virus.

FH Key Location/Qualifiers

FT Protein 628..993

FT /note= "claimed; X denotes translated stop codons  
 and unspecified triplets"

XX EPI38704-A.

XX 24-APR-1985.

XX 09-OCT-1984; 84EP-0402025.

XX 02-MAR-1984; 84US-0585942.

XX 14-OCT-1983; 83US-0541836.

XX (MERI ) MERCK &amp; CO INC.

XX Hughes JV, Scolnick EM, Tomassini JE;

XX WPI; 1985-100818/17.

XX N-PSDB; N50274.

PT New hepatitis A virus surface protein - useful for binding to  
 neutralising antibodies to the virus

PS Disclosure; Page 17-23; 49pp; English.

XX VPI is isolated by solubilisation of the intact virus in an aq.  
 CC antonic surfactant and a reducing agent. The viral proteins are sepd.  
 CC and the protein of molecular wt. 33000 daltons is sepd.

SQ Sequence 993 AA;

Query Match 96.0%; Score 97; DB 6; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-08;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHEIMKF 19  
 |||||  
 Db 970 fsqakislfyteeheimkf 988

RESULT 14

W76539  
 ID W76539 standard; Protein; 20 AA.

XX W76539;

XX 11-DEC-1998 (first entry)

XX Plasmid pGEM2-HAV-delta2BC protein fragment.

KW Picornavirus; structural protein; P1-2A; P1; P3 region; HAV;  
 KW hepatitis A virus; viral replication; empty viral particle.

XX Synthetic.

XX WO9844122-A1.

XX 08-OCT-1998.

XX 26-MAR-1998; 98WO-DE00879.

XX 27-MAR-1997; 97DE-1012899.

XX (NOVE-) NOVEMBER NOVUS MEDICATUS BERTLING GES MO.

XX Probst C;

XX WPI; 1998-557117/47.

XX N-PSDB; V61866.

PT Recombinant production of picorna virus - by co-expressing  
 structural protein precursors P1 and P3, in cis or trans,  
 particularly to produce hepatitis A particles for use in vaccines

PS Disclosure; Fig 3; 25pp; German.

XX This sequence represents a fragment of the plasmid pGEM2-HAV-delta2BC  
 CC which is used in a novel method which results in the recombinant  
 CC production of picornavirus particles, their precursors or partial  
 CC derivatives and involves co-expressing structural protein precursor  
 CC molecules P1-2A or P1, with the complete P3 region (3ABCD), in cis or in  
 CC trans. This invention has particular use with hepatitis A virus (HAV)  
 CC particles. This method, which does not involve expressing the P2B or 2C  
 CC proteins, is more efficient and excludes the possibility of viral  
 CC replication. Using the complete P3 region, rather than just 3C and 3D,  
 CC significantly improves production of empty viral particles.

SQ Sequence 20 AA;

Query Match 49.0%; Score 49.5; DB 19; Length 20;  
 Best Local Similarity 75.0%; Pred. No. 0.086;  
 Matches 12; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 3 QAKISLFYTEEHEIMK 18

Db 1 qanislfytee-efme 15

RESULT 15

Y27340  
 ID Y27340 standard; protein; 518 AA.

XX Y27340;

XX 15-NOV-1999 (first entry)

XX Group B Streptococcus (GBS) antigen (clone 1).

XX Group B Streptococcus; GBS; antigen; vaccine; Streptococcus infection;  
 KW sepsis; meningitis; pneumonia; immunocompromise; diabetes; liver disease;  
 KW cancer; veterinary; mastitis.

XX Streptococcus sp.

XX OS  
 XX W09942588-A2.

```

XX 26-AUG-1999.
PD
XX
XX 17-FEB-1999; 99WO-CA00114.
PF
XX 20-FEB-1998; 98US-0075425.
PR
XX
XX (BIOC-) BIOCHEM VACCINS INC.
PA
XX
XX Boyer M, Brodeur BR, Charlebois I, Hamel J, Martin D;
PI Rioux C;
PI
XX
XX WPI; 1999-540309/45.
DR
DR N-PSDB; X91103.
XX
XX Novel group B Streptococcus antigens - useful as vaccine
PT compositions for prophylaxis or therapy of Streptococcus infections
PT
XX
XX Claim 26; Fig 1F; 154pp; English.
PS
XX
XX The invention provides Group B Streptococcus (GBS) antigens (Y27336-370)
CC and nucleic acids (X91103-X91111) encoding the antigens. The GBS antigens
CC can be recombinantly expressed using standard recombinant methodology.
CC The GBS antigens of the invention can be used as vaccine components for
CC the treatment or prophylaxis of diseases and symptoms mediated by
CC Streptococcus infection, especially group A Streptococcus (S. pyogenes),
CC GBS or S. agalactiae, S. dysgalactiae, S. uberis, S. noxia, as well as
CC Staphylococcus aureus. The vaccines are administered to those individuals
CC at risk of GBS infection, particularly pregnant women and infants for
CC sepsis, meningitis, and pneumonia, as well as immunocompromised
CC individuals, such as those with diabetes, liver disease or cancer. The
CC vaccines also have veterinary applications, such as for the treatment of
CC mastitis in cattle. The present sequence represents a GBS antigen of the
CC invention.
XX
XX Sequence 518 AA;
SQ

```

Query Match 46.5%; Score 47; DB 20; Length 518;  
 Best Local Similarity 45.0%; Pred. No. 8.2;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FSQAKISLFTTEHEIMKFS 20  
 | | | | | | | | | |  
 Db 371 fcqapialvknahhevavafa 390

Search completed: April 24, 2001, 16:23:27  
 Job time: 145 sec



Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	1091	6	5516630-2	Patent No. 5516630
2	101	100.0	2227	4	US-08-475-886-2	Sequence 2, Appli
3	101	100.0	2227	4	US-08-475-886-4	Sequence 4, Appli
4	101	100.0	2227	4	US-08-475-886-6	Sequence 6, Appli
5	101	100.0	2227	4	US-08-397-232-2	Sequence 2, Appli
6	101	100.0	2227	4	US-08-397-232-4	Sequence 4, Appli
7	41	40.6	2431	1	US-07-920-281C-2	Sequence 2, Appli
8	41	40.6	2431	1	US-08-466-277-2	Sequence 2, Appli
9	39	38.6	289	5	PCT-US92-00282-11	Sequence 11, Appl
10	39	38.6	860	1	US-08-092-817-4	Sequence 4, Appli
11	39	38.6	1410	2	US-08-470-058-4	Sequence 4, Appli
12	39	38.6	1410	3	US-09-037-188-4	Sequence 4, Appli
13	38	37.6	265	1	US-07-958-551-2	Sequence 2, Appli
14	38	37.6	265	1	US-08-129-610-7	Sequence 7, Appli
15	38	37.6	265	1	US-08-129-609A-7	Sequence 7, Appli
16	38	37.6	265	1	US-08-455-313-7	Sequence 7, Appli
17	38	37.6	265	1	US-08-475-924-2	Sequence 2, Appli
18	38	37.6	265	2	US-08-657-579A-2	Sequence 2, Appli
19	38	37.6	265	4	US-09-224-025-7	Sequence 7, Appli
20	38	37.6	265	5	PCT-US94-07887-7	Sequence 7, Appli
21	38	37.6	707	3	US-08-772-270A-4	Sequence 4, Appli
22	37	36.6	198	1	US-08-097-827-2	Sequence 2, Appli
23	37	36.6	198	1	US-08-494-574-2	Sequence 2, Appli
24	37	36.6	420	3	US-08-685-871-58	Sequence 12, Appl
25	37	36.6	711	3	US-08-772-270A-12	Sequence 58, Appl
26	37	36.6	1354	3	US-08-685-871-2	Sequence 2, Appli
27	37	36.6	1388	2	US-08-685-576-1	Sequence 1, Appli

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20
   |||||||
Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 3
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475.886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1993-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20
   |||||||
Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 4
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475.886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1993-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20
   |||||||
Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 5
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20
   |||||||
Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 6
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4
```

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.2e+08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
DB 834 FSOAKISLFYTEHEIMKFS 853

## RESULT 7

US-07-920-281C-2  
; Sequence 2, Application US/07920281C  
; Patent No. 5739026  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; TITLE OF INVENTION: Alphaviruses  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/920,281C  
; FILING DATE: 13-AUG-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-920-281C-2

Query Match 40.6%; Score 41; DB 1; Length 2431;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEEH 14  
DB 1072 FSAPKVSLEYENN 1085

## RESULT 8

US-08-466-277-2  
; Sequence 2, Application US/08466277  
; Patent No. 6190666  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; TITLE OF INVENTION: Alphaviruses  
; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,277  
; FILING DATE: 06-Jun-1995  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/920,281  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-08-466-277-2  
  
Query Match 40.6%; Score 41; DB 4; Length 2431;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 FSOAKISLFYTEEH 14  
DB 1072 FSAPKVSLEYENN 1085  
  
RESULT 9  
PCT-US92-00282-11  
; Sequence 11, Application PC/TUS9200282  
; GENERAL INFORMATION:  
; APPLICANT: OWENS, IDA S.  
; APPLICANT: RITTER, JOSEPH K.  
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION  
; TITLE OF INVENTION: THEREIN.  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; STREET: 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00282  
; FILING DATE: 19920110  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.

REGISTRATION NUMBER: 26581  
REFERENCE/DOCKET NUMBER: 91532-PCT  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 289 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US92-00282-11

Query Match 38.6%; Score 39; DB 5; Length 289;  
Best Local Similarity 58.3%; Pred. No. 33;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 FYTEEHKMS 20  
||:| ||:| |  
Db 100 FFTEHLLKMS 111

RESULT 10  
US-08-092-817-4  
; Sequence 4, Application US/08092817  
; Patent No. 5496926  
; GENERAL INFORMATION:  
; APPLICANT: RUBINSTEIN, Menachem  
; APPLICANT: NOVICK, Daniela  
; APPLICANT: TAL, Nathan  
; TITLE OF INVENTION: SOLUBLE LDL RECEPTOR, ITS PRODUCTION AND  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/092,817  
; FILING DATE: 19-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 100696  
; FILING DATE: 19-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 102915  
; FILING DATE: 23-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: TOWNSEND, G. Kevin  
; REGISTRATION NUMBER: 34, 033  
; REFERENCE/DOCKET NUMBER: RUBINSTEIN-5A  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 860 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-092-817-4

Query Match 38.6%; Score 39; DB 1; Length 860;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 LFYTEEHKMS 20  
||:| ||:| |  
Db 401 LFFTHRHEVRKMT 413

RESULT 11  
US-08-470-058-4  
; Sequence 4, Application US/08470058  
; Patent No. 5817789  
; GENERAL INFORMATION:  
; APPLICANT: Heartlein, Michael W.  
; APPLICANT: Lemontt, Jeffrey F.  
; TITLE OF INVENTION: Chimeric Proteins For Use in Transport  
; TITLE OF INVENTION: of a Selected Substance Into Cells  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,058  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: TKT93-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1410 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-470-058-4

Query Match 38.6%; Score 39; DB 2; Length 1410;  
Best Local Similarity 46.2%; Pred. No. 1.9e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 LFYTEEHKMS 20  
||:| ||:| |  
Db 401 LFFTHRHEVRKMT 413

RESULT 12  
US-09-037-188-4  
; Sequence 4, Application US/09037188  
; Patent No. 6027921  
; GENERAL INFORMATION:  
; APPLICANT: Heartlein, Michael W.  
; APPLICANT: Lemontt, Jeffrey F.  
; APPLICANT: Concino, Michael F.  
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR USE  
; TITLE OF INVENTION: IN TRANSPORT OF A SELECTED SUBSTANCE INTO CELLS  
US-09-037-188-4

NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/037/188  
FILING DATE: 02-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Ph.D., J.D., Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 07236/009002  
TELEPHONE: 617/542-8906  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1410 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-037-188-4

Query Match 38.6%; Score 39; DB 3; Length 1410;  
Best Local Similarity 46.2%; Pred. No. 1.9e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 LFVTEPEHEIMKFS 20  
11:1 11:1  
DB 401 LFTNRHEVRKMT 413

RESULT 13  
US-07-958-551-2  
Sequence 2, Application US/07958551  
Patent No. 5302387  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel M.  
APPLICANT: Kennedy, Keith M.  
APPLICANT: Randall, John Brooks  
APPLICANT: Brower, David Orlin  
TITLE OF INVENTION: Bacillus thuringiensis Isolates Active Against  
TITLE OF INVENTION: Cockroaches and Genes Encoding Cockroach-Active  
TITLE OF INVENTION: Toxins  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/958,551  
FILING DATE: October 19, 1992

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/788,654  
FILING DATE: No. 5302387ember 6, 1991  
NAME:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/J 101.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleoensis  
INDIVIDUAL ISOLATE: PS201T6  
IMMEDIATE SOURCE:  
LIBRARY: LambdaGem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-07-958-551-2

Query Match 37.6%; Score 38; DB 1; Length 265;  
Best Local Similarity 58.3%; Pred. No. 44;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 KISLFYTEHEI 16  
1:1:11111  
DB 2 KESIYNEENEI 13

RESULT 14  
US-08-129-610-7  
Sequence 7, Application US/08129610  
Patent No. 5436002  
GENERAL INFORMATION:  
APPLICANT: Jewel Payne  
APPLICANT: Kenneth Narva  
APPLICANT: Kendrick Akira Uyeda  
APPLICANT: Christine Julie Stalder  
APPLICANT: Tracy Ellis Michaels  
TITLE OF INVENTION: No. 5436002el Bacillus thuringiensis Isolates and Toxins  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/129,610  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,199  
FILING DATE: 15-JUL-1993  
NAME:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/977,350  
FILING DATE: 17-NOV-1992  
NAME:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/746,751  
FILING DATE: 21-AUG-1991  
NAME:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/708,266  
FILING DATE: 28-MAY-1991  
NAME:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,399  
FILING DATE: 29-JAN-1991  
NAME:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: MA55CCD.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleoensis  
INDIVIDUAL ISOLATE: PS201T6  
IMMEDIATE SOURCE:  
LIBRARY: LambdaGem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-08-129-610-7

Query Match 37.6%; Score 38; DB 1; Length 265;  
Best Local Similarity 58.3%; Pred. No. 44;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KISLFYTEEHEI 16  
| |:::| |:::|  
Db 2 KESIYNEENEI 13

RESULT 15

US-08-129-609A-7  
; Sequence 7, Application US/08129609A  
; Patent No. 5489432  
; GENERAL INFORMATION:  
; APPLICANT: Jewel Payne  
; APPLICANT: M. Keith Kennedy  
; APPLICANT: John Brookes Randall  
; APPLICANT: David Orlin Brower  
; APPLICANT: H. Ernest Schnepf  
; TITLE OF INVENTION: Bacillus thuringiensis Isolates Active  
; TITLE OF INVENTION: Against Cockroaches and Genes Encoding Cockroach-Active Toxins  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/129,609A  
FILING DATE: 30-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/958,551  
FILING DATE: 19-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/788,654  
FILING DATE: 6-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/J 101.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
STRAIN: neoleoensis  
INDIVIDUAL ISOLATE: PS201T6  
IMMEDIATE SOURCE:  
LIBRARY: LambdaGem (TM)-11 Library of Kenneth E. Narva  
CLONE: 201T635  
US-08-129-609A-7

Query Match 37.6%; Score 38; DB 1; Length 265;  
Best Local Similarity 58.3%; Pred. No. 44;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 KISLFYTEEHEI 16  
| |:::| |:::|  
Db 2 KESIYNEENEI 13

Search completed: April 24, 2001, 16:35:04  
Job time: 535 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:25:54 ; Search time 73.14 Seconds  
(without alignments)  
18.792 Million cell updates/sec

Title: US-09-171-432A-43  
Perfect score: 101

Sequence: 1 FSOAKISLFYTEHEIMKFS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_67:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	1358	2 A03905	genome polyprotein
2	101	100.0	2227	1 GNNYHM	genome polyprotein
3	101	100.0	2227	1 GNNYHR	genome polyprotein
4	101	100.0	2227	1 GNNYMK	genome polyprotein
5	101	100.0	2227	1 GNNYHB	genome polyprotein
6	97	96.0	852	1 GNNYHA	genome polyprotein
7	88	87.1	2230	1 GNNYSA	genome polyprotein
8	47	46.5	378	2 F64300	formate dehydrogen
9	43	42.6	325	2 H71869	hypothetical prote
10	43	42.6	410	2 T22413	hypothetical prote
11	42	41.6	187	2 T18844	hypothetical prote
12	42	41.6	545	2 S10842	gene ND5 intron 2
13	41.5	41.1	288	2 S64243	hypothetical prote
14	41	40.6	144	2 E75128	glycine cleavage s
15	41	40.6	209	2 S65300	hypothetical prote
16	41	40.6	304	2 C72324	homoserine O-succi
17	41	40.6	483	2 G84113	hypothetical prote
18	41	40.6	557	2 S58164	glucose-6-phosphat
19	41	40.6	820	2 T51510	hypothetical prote
20	41	40.6	1040	1 A38306	alpha-mannosidase
21	41	40.6	2431	1 MNVVSF	nonstructural poly
22	40.5	40.1	251	2 T13104	minor tail protein
23	40.5	40.1	419	2 T25577	hypothetical prote
24	40	39.6	142	2 C70059	hypothetical prote
25	40	39.6	203	2 S63633	hypothetical prote
26	40	39.6	256	2 T45895	hypothetical prote
27	40	39.6	333	2 T33866	hypothetical prote
28	40	39.6	450	2 C82204	heat shock protein
29	40	39.6	523	2 D83631	probable sulfate t

30	39.5	39.1	109	2 S74054	hypothetical prote
31	39.5	39.1	260	2 B71353	probable D,D-carbo
32	39.5	39.1	410	2 T23264	hypothetical prote
33	39	38.6	129	2 F72075	hypothetical prote
34	39	38.6	145	2 F83247	hypothetical prote
35	39	38.6	152	1 C69202	conserved hypotet
36	39	38.6	167	2 F83845	hypothetical prote
37	39	38.6	171	2 T18167	hypothetical prote
38	39	38.6	289	2 B42586	glucuronosyltransf
39	39	38.6	306	1 BVBXK6	MAK16 protein - ye
40	39	38.6	319	2 F81319	ADP-heptose-LPS h
41	39	38.6	380	1 NPBY	anthranilate phosph
42	39	38.6	382	2 T14186	hypothetical prote
43	39	38.6	398	2 F70353	RNA polymerase sig
44	39	38.6	433	2 A72090	hypothetical prote
45	39	38.6	435	2 A81580	hypothetical prote

ALIGNMENTS

RESULT 1

A03905

genome polyprotein (version 2) - human hepatitis A virus (fragments)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; prote

C:Species: human hepatitis A virus

C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996

C:Accession: A03905

R:Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feins

Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985

A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and R

A:Reference number: A03905; MUID:85166289

A:Accession: A03905

A:Molecule type: genomic RNA

A:Residues: 1-1358 <BAR>

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; polyprotein

F:1-245/Product: coat protein 1A #status predicted <C1A>

F:246-491/Product: coat protein 1B #status predicted <C1B>

F:492-836/Product: coat protein 1C #status predicted <C1C>

F:837-854/Product: core protein 2A (fragment) #status predicted <C2A>

F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 101; DB 2; Length 1358;

Best Local Similarity 100.0%; Pred. No. 1.2e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20

|||||

Db 834 FSOAKISLFYTEHEIMKFS 853

RESULT 2

GNNYHM

genome polyprotein - human hepatitis A virus (strain HM-175, wild type)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core

B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D

C:Species: human hepatitis A virus

A:Note: host Homo sapiens (man)

C>Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C:Accession: A25981

R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.

J. Virol. 61, 50-59, 1987

A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with

A:Reference number: A25981; MUID:87061253

A:Accession: A25981

A:Molecule type: genomic RNA

A:Residues: 1-2227 <COH>

A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr

F:1-23/Product: coat protein 1A #status predicted <VP4>

A:Reference number: A94149; MUID:87175701  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M16632; NID:g329594; PIDN:AAA5471.1; PID:g329595  
C:Note: submitted to GenBank, August 1987  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.le-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFTTEHEIMKFS 20  
|||||  
Db 834 FSOAKISLFTTEHEIMKFS 853

RESULT 5  
GNNYHB  
genome polyprotein - human hepatitis A virus (strain MBB)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
Vpg; protein 3A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: JS0303  
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wissel, T.; Klehn, R.; Wimmer, E.; Deinhard  
Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (iso  
A:Reference number: JS0303; MUID:88045071  
A:Accession: JS0303  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <PAU>  
A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; h  
F:1-23/Product: coat protein 1A #status predicted <VP1>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP4>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1108/Product: core protein 2B #status predicted <P2B>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1496/Product: protein 3A #status predicted <P3A>  
F:1497-1519/Product: genome-linked protein vpg #status predicted <VPG>  
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.le-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFTTEHEIMKFS 20  
|||||  
Db 834 FSOAKISLFTTEHEIMKFS 853

RESULT 6  
GNNYHA

genome polyprotein - human hepatitis A virus (strain CR326) (fragment)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 2A  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03904  
R:Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, S.W.  
J. Virol. 54, 247-255, 1985  
A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.  
A:Reference number: A03904; MUID:85185648  
A:Accession: A03904  
A:Molecule type: genomic RNA  
A:Residues: 1-852 <LIN>  
A:Cross-references: EMBL:M10033; NID:G329592; PIDN:AAA45470.1; PID:G329593  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <CIA>  
F:246-491/Product: coat protein 1B #status predicted <CIB>  
F:492-836/Product: coat protein 1C #status predicted <CIC>  
F:837-852/Product: coat protein 2A (fragment) #status predicted <C2A>

Query Match 96.0%; Score 97; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 3.3e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 19  
DB 834 FSOAKISLFYTEHEIMKFS 852  
|||||

RESULT 7  
GNVSYA  
genome polyprotein - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
C:Species: simian hepatitis A virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C:Accession: A30470; S04885; S03965  
R:Tzarev, S.A.  
submitted to JPIID, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:G222597; PIDN:BAA00766.1; PID:G222598  
R:Tzarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A:Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure an  
A:Reference number: JQ1080; MUID:91311420  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhiko  
submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: genomic RNA  
A:Residues: 1750-2164 <BAL1>  
A:Cross-references: EMBL:X15461; NID:G61971; PIDN:CAA33490.1; PID:G930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhiko  
FEBS Lett. 247, 425-428, 1989  
A:Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A:Reference number: S03965; MUID:89232168  
A:Accession: S03965  
A:Molecule type: genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <CIA>  
F:28-249/Product: coat protein 1B #status predicted <CIB>  
F:250-495/Product: coat protein 1C #status predicted <CIC>  
F:496-795/Product: coat protein 1D #status predicted <CID>  
F:796-984/Product: coat protein 2A #status predicted <C2A>

F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: core protein 3A #status predicted <P3A>  
F:1499-1521/Product: core protein 3B #status predicted <P3B>  
F:1522-1741/Product: core protein 3C #status predicted <P3C>  
F:1742-2230/Product: core protein 3D #status predicted <P3D>

Query Match 87.1%; Score 88; DB 1; Length 2230;  
Best Local Similarity 85.0%; Pred. No. 3.2e-06;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEIMKFS 20  
DB 838 FSOAKISLFYTEHEIMKFS 857  
|||||

RESULT 8  
F64300  
formate dehydrogenase (EC 1.2.1.2) alpha chain - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: F64300  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak  
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc  
A:Reference number: A64300; MUID:96337999  
A:Accession: F64300  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-378 <BUL>  
A:Cross-references: GB:U67459; GB:L77117; NID:G2826236; PIDN:AAB97987.1; PID:gl498763  
C:Genetics:  
A:Map position: REV8474-7338  
C:Keywords: oxidoreductase

Query Match 46.5%; Score 47; DB 2; Length 378;  
Best Local Similarity 47.1%; Pred. No. 3.3;  
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 4 AKISLFYTEHEIMKFS 20  
DB 174 SKVTIFTEKEILKLN 190  
|||||

RESULT 9  
H71869  
hypothetical protein jhp0940 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
C:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 04-Mar-2000  
C:Accession: H71869  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p  
A:Reference number: A71800; MUID:99120557  
A:Accession: H71869  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-325 <ARN>  
A:Cross-references: GB:AE001523; GB:AE001439; NID:G4155513; PIDN:AAD06516.1; PID:G415  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0940  
C:Superfamily: Helicobacter pylori hypothetical protein jhp0940

Query Match 42.6%; Score 43; DB 2; Length 325;

Best Local Similarity 60.0%; Pred. No. 13;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 KISLFYTEEHEIMKF 19  
| | | | | | : | | |  
Db 24 KISLFYNELMYVKF 38

RESULT 10  
T22413  
hypothetical protein F49C12.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22413  
R:Gardner, A.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19562  
A:Accession: T22413  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-410 <WIL>  
A:Cross-references: EMBL:Z68227; PIDN:CAA92512.1; GSPDB:GN00022; CESP:F49C12.8  
A:Experimental source: clone F49C12  
C:Genetics:  
A:Gene: CESP:F49C12.8  
A:Map position: 4  
A:Introns: 21/3; 302/3

Query Match 42.6%; Score 43; DB 2; Length 410;  
Best Local Similarity 36.8%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHEIMKF 19  
| : : : | | : | | |  
Db 156 FAMIRVGLFLDHLINKF 174

RESULT 11  
T18844  
hypothetical protein C01H6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T18844  
R:Berk, M.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19030  
A:Accession: T18844  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-187 <WIL>  
A:Cross-references: EMBL:Z71258; PIDN:CAA95778.1; GSPDB:GN00019; CESP:C01H6.3  
A:Experimental source: clone C01H6  
C:Genetics:  
A:Gene: CESP:C01H6.3  
A:Map position: 1  
A:Introns: 42/3; 82/1; 109/1  
C:Superfamily: Caenorhabditis elegans hypothetical protein C01H6.3

Query Match 41.6%; Score 42; DB 2; Length 187;  
Best Local Similarity 36.8%; Pred. No. 10;  
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHEIMKF 19  
| | : | | : | : | |  
Db 161 FCNARISLQHAQSHQCVRF 179

RESULT 12  
SI0842  
gene ND5 intron 2 protein - Neurospora crassa mitochondrion  
C:Species: mitochondrion Neurospora crassa

Best Local Similarity 60.0%; Pred. No. 13;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 KISLFYTEEHEIMKF 19  
| | | | | | : | | |  
Db 24 KISLFYNELMYVKF 38

RESULT 10  
T22413  
hypothetical protein F49C12.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22413  
R:Gardner, A.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19562  
A:Accession: T22413  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-410 <WIL>  
A:Cross-references: EMBL:Z68227; PIDN:CAA92512.1; GSPDB:GN00022; CESP:F49C12.8  
A:Experimental source: clone F49C12  
C:Genetics:  
A:Gene: CESP:F49C12.8  
A:Map position: 4  
A:Introns: 21/3; 302/3

Query Match 42.6%; Score 43; DB 2; Length 410;  
Best Local Similarity 36.8%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHEIMKF 19  
| : : : | | : | | |  
Db 156 FAMIRVGLFLDHLINKF 174

RESULT 11  
T18844  
hypothetical protein C01H6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T18844  
R:Berk, M.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19030  
A:Accession: T18844  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-187 <WIL>  
A:Cross-references: EMBL:Z71258; PIDN:CAA95778.1; GSPDB:GN00019; CESP:C01H6.3  
A:Experimental source: clone C01H6  
C:Genetics:  
A:Gene: CESP:C01H6.3  
A:Map position: 1  
A:Introns: 42/3; 82/1; 109/1  
C:Superfamily: Caenorhabditis elegans hypothetical protein C01H6.3

Query Match 41.6%; Score 42; DB 2; Length 187;  
Best Local Similarity 36.8%; Pred. No. 10;  
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHEIMKF 19  
| | : | | : | : | |  
Db 161 FCNARISLQHAQSHQCVRF 179

RESULT 12  
SI0842  
gene ND5 intron 2 protein - Neurospora crassa mitochondrion  
C:Species: mitochondrion Neurospora crassa

C:Date: 28-May-1993 #sequence\_revision 28-May-1993 #text\_change 21-Jul-2000  
C:Accession: SI0842; SI0081  
R:Nelson, M.A.; Macino, G.  
Mol. Gen. Genet. 206, 307-317, 1987  
A:Title: Structure and expression of the overlapping ND4L and ND5 genes of Neurospora  
A:Reference number: S07320; MUID:87228330  
A:Accession: SI0842  
A:Molecule type: DNA  
A:Residues: 1-545 <NEL>  
A:Cross-references: EMBL:X05115  
R:Almasan, A.; Mishra, N.C.  
Genetics 120, 935-945, 1988  
A:Title: Molecular characterization of the mitochondrial DNA of a new stopper mutant  
A:Reference number: SI0081; MUID:89137935  
A:Accession: SI0081  
A:Molecule type: DNA  
A:Residues: 457-545 <ALM>  
A:Cross-references: EMBL:X14681; NID:gl4021; PIDN:CAA32812.1; PID:gl4022  
A:Experimental source: strain Lindegren-Rockefeller RL3-8A  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC3  
A:Introns: 108/3  
C:Keywords: mitochondrion

Query Match 41.6%; Score 42; DB 2; Length 545;  
Best Local Similarity 41.2%; Pred. No. 34;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 FSQAKISLFYTEEHEIM 17  
| | : | | : | | : |  
Db 241 FSRAFLKLYMQEHPVL 257

RESULT 13  
S64243  
hypothetical protein YGL221c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein Gl020  
C:Species: Saccharomyces cerevisiae  
C:Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 20-Jun-2000  
C:Accession: S64243  
R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64071  
A:Accession: S64243  
A:Molecule type: DNA  
A:Residues: 1-288 <RIE>  
A:Cross-references: EMBL:Z72743; NID:gl322868; PIDN:CAA96937.1; PID:gl322869; MIPS:YG  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:NIF3  
A:Cross-references: SGD:S0003189; MIPS:YGL221c  
A:Map position: 7L  
C:Superfamily: conserved hypothetical protein YGL144c

Query Match 41.1%; Score 41.5; DB 2; Length 288;  
Best Local Similarity 41.7%; Pred. No. 20;  
Matches 10; Conservative 5; Mismatches 4; Indels 5; Gaps 2;

QY 1 FSQAK--ISLFYTEE---HEIMKF 19  
| | | : | | | | | : | : |  
Db 217 FKQLKEDVDLYTGTGMSHHEVLKW 240

RESULT 14  
E75128  
glycine cleavage system protein h PAB0559 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: E75128  
R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: E75128  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-144 <KAW>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49742.1; PID:g545825  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: gcsH4; PAB0559  
C:Superfamily: glycine cleavage system protein H; lipoyl/biotin-binding homology  
F:41-115/Domain: lipoyl/biotin-binding homology <LPB>

Query Match 40.6%; Score 41; DB 2; Length 144;  
Best Local Similarity 53.8%; Pred. No. 11;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 QAKISLFYTEEHE 15  
: | | | | | | | | | |  
Db 18 KVKEGLYTKHE 30

RESULT 15  
S65300  
hypothetical protein YPL267w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein P0360  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Dec-1994 #sequence\_revision 31-May-1996 #text\_change 20-Jun-2000  
A:Accession: S65300; S65321  
R:Duesterhoeft, A.; Floeth, M.; Fritz, M.; Hilbert, H.; Moestl, D.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S65292  
A:Accession: S65300  
A:Molecule type: DNA  
A:Residues: 1-209 <DUE>  
A:Cross-references: EMBL:Z73623; NID:gl370550; PID:gl370551; MIPS:YPL267w  
A:Experimental source: strain S288C (AB972)  
R:Delius, H.; Hebling, U.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64967  
A:Accession: S65321  
A:Molecule type: DNA  
A:Residues: 1-209 <DEL>  
A:Cross-references: EMBL:Z73623; NID:gl370550; PID:gl370551; MIPS:YPL267w  
A:Experimental source: strain S288C (AB972)  
C:Genetics:  
A:Map position: 16L  
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YPL267w

Query Match 40.6%; Score 41; DB 2; Length 209;  
Best Local Similarity 50.0%; Pred. No. 17;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 4 AKISLFYTEEHE 15  
: | | | | | | | | | |  
Db 166 SKISIFFTSKHQ 177

Search completed: April 24, 2001, 16:25:57  
Job time: 291 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:34:00 ; Search time 38.74 Seconds  
(without alignments)  
17.685 Million cell updates/sec

Title: US-09-171-432A-43  
Perfect score: 101  
Sequence: 1 FSOAKISLFYTEHEIMKFS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	2227	1 POLG_HPAVH	P08617 hepatitis a
2	101	100.0	2227	1 POLG_HPAVL	P08441 hepatitis a
3	101	100.0	2227	1 POLG_HPAVM	P13901 hepatitis a
4	97	96.0	852	1 POLG_HPAVC	P06442 hepatitis a
5	93	92.1	2226	1 POLG_HPAV2	P26580 hepatitis a
6	93	92.1	2226	1 POLG_HPAV4	P26581 hepatitis a
7	93	92.1	2226	1 POLG_HPAV8	P26582 hepatitis a
8	88	87.1	2230	1 POLG_HPAVS	P14553 simian hepa
9	47	46.5	378	1 FDHA_METUA	Q60314 methanococc
10	43	42.6	410	1 YE28_CAEEL	Q20585 caenorhabdi
11	43	42.6	722	1 YC06_KLEPN	Q48452 klebsiella
12	41.5	41.1	288	1 NIF3_YEAST	P53081 saccharomyc
13	41	40.6	138	1 GCSH_PYRAB	Q9Y091 pyrococcus
14	41	40.6	304	1 META_THEMEA	Q9WZY3 thermotoga
15	41	40.6	557	1 GFLI_ACICA	Q59088 acinetobact
16	41	40.6	1040	1 MANI_RAT	P21139 rattus norv
17	41	40.6	2431	1 POLN_SFV	P08411 semliki for
18	40	39.6	418	1 B2AR_BOVIN	Q28044 bos taurus
19	39	38.6	306	1 MK16_YEAST	P10962 saccharomyc
20	39	38.6	380	1 TRPD_YEAST	P07285 saccharomyc
21	39	38.6	534	1 UDI15_HUMAN	P35504 homo sapien
22	39	38.6	634	1 YCX3_EUGGR	P31916 euglena gra
23	39	38.6	707	1 HLY2_ECOLI	P10089 escherichia
24	39	38.6	707	1 HLYB_ECOLI	P08716 escherichia
25	39	38.6	707	1 HLYB_PROVU	P11599 proteus vul
26	39	38.6	837	1 LDLR_RABIT	P20063 oryctolagus
27	39	38.6	860	1 LDLR_HUMAN	P01130 homo sapien
28	39	38.6	892	1 LDL2_XENLA	Q99088 xenopus lae
29	39	38.6	909	1 LDL1_XENLA	Q99087 xenopus lae
30	38	37.6	48	1 Y234_HAEIN	P43968 haemophilus
31	38	37.6	138	1 GCSH_PYRHO	O59049 pyrococcus
32	38	37.6	265	1 CXBA_SALTY	Q45790 bacillus th
33	38	37.6	317	1 RFAC_SALTY	P26469 salmonella

RESULT 1

POLG\_HPAVH STANDARD; PRT; 2227 AA.  
AC P08617; P06443; Q81082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain HM-175).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12098;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WILD TYPE;  
RX MEDLINE=87061253; PubMed=3023706;  
RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,  
RA Baroudy B.M.;  
RT "Complete nucleotide sequence of wild-type hepatitis A virus:  
RT comparison with different strains of hepatitis A virus and other  
RT picornaviruses.";  
RL J. Virol. 61:50-59(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATTENUATED;  
RX MEDLINE=87175701; PubMed=3031686;  
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,  
RA Purcell R.H.;  
RT "Complete nucleotide sequence of an attenuated hepatitis A virus:  
RT comparison with wild-type virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
RN [3]  
RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
RX MEDLINE=85166289; PubMed=2984684;  
RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,  
RA Purcell R.H., Feinstone S.M.;  
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid  
RT proteins and RNA polymerase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -!- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED  
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3  
CC -!- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
CC SHOWN.

ALIGNMENTS

34 38 37.6 319 1 RFAC\_ECOLI P24173 escherichia  
35 38 37.6 377 1 MTC3\_CHVNI P10835 chlorella v  
36 38 37.6 396 1 REPA\_BAGSU P13962 bacillus v  
37 38 37.6 400 1 ACH1\_BOMMO Q03383 bombyx mori  
38 38 37.6 407 1 ABRH\_LOCMI P32122 locusta mig  
39 38 37.6 448 1 RADA\_HELPY P56148 helicobacte  
40 38 37.6 460 1 DB80\_DROME Q61305 drosophila  
41 38 37.6 492 1 PD11\_SCHPO Q10057 schizosacch  
42 38 37.6 533 1 MCP4\_ECOLI P07018 escherichia  
43 38 37.6 707 1 P71B\_ACTPL P26760 actinobacil  
44 38 37.6 784 1 KL68\_DROME P46867 drosophila  
45 38 37.6 1660 1 VIT6\_OSCBR Q94637 oscheilus br

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; M14114; AAA45475.1; -
CC EMBL; M14707; AAA45465.1; -
CC EMBL; M14707; AAA45466.1; ALT_INIT.
CC EMBL; M16632; AAA45471.1; -
CC PIR; A25981; GNNYHM.
CC PIR; A25914; GNNYMK.
CC PIR; A03905; A03905.
CC MEROPS; C03.005; -
CC InterPro; IPR001205; -
CC Pfam; PF00910; RNA_dep_RNA_pol; 1.
CC Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 836
FT CHAIN 837 980
FT CHAIN 981 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1496
FT CHAIN 1497 1519
FT CHAIN 1520 1738
FT CHAIN 1739 2227
FT VARIANT 77 77
FT VARIANT 764 764
FT VARIANT 821 821
FT VARIANT 1052 1052
FT VARIANT 1062 1062
FT VARIANT 1118 1118
FT VARIANT 1151 1151
FT VARIANT 1163 1163
FT VARIANT 1277 1277
FT VARIANT 1500 1500
FT VARIANT 1805 1805
FT VARIANT 1930 1930
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AE740A6 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEEHEIMKFS 20
| | | | | | | | | | | | | | | | | | | | | |
Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 2
POLG_HPAVL STANDARD; PRT; 2227 AA.
AC P06441;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain LA).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12099;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85190549; PubMed=2986127;
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,
RA Merryweather J., van Nest G., Dina D.;
RT "Primary structure and gene organization of human hepatitis A virus";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).

-!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
-!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

-----

CC EMBL: M10033; AAA45470.1; -  
CC PIR: A03904; GNNYHA.  
DR Polyprotein; Coat protein; Core protein.  
KW Polyprotein; Coat protein; Core protein.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).  
FT CHAIN 837 >852 CORE PROTEIN P2A.  
FT NON\_TER 852 852  
FT SEQUENCE 852 AA; 955663 MW: 73D3ED0AD532820E CRC64;  
SO

```

Query Match          96.0%; Score 97; DB 1: Length 852;
Best Local Similarity 100.0%; Pred. No. 1.le-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  FSQAKISLFYTEEHEIMKF 19
        |||||
Db       834  FSQAKISLFYTEEHEIMKF 852

```

DE	P3D (EC 2.7.7.48)].	
OS	Hepatitis A virus (strain 24a).	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;	
OC	Hepatovirus.	
OX	NCBI_TaxID=12094;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=91162758; PubMed=1705995;	
RA	Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,	
RA	Cromean T., Jansen R.W.;	
RT	"Antigenic and genetic variation in cytopathic hepatitis A virus	
RT	variants arising during persistent infection: evidence for genetic	
RT	recombination.";	
RL	J. Virol. 65:2056-2065(1991).	
CC	-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,	
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,	
CC	VP3, AND VP4.	
CC	-1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.	
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.	

CC. modified and this statement is not removed. Usage by and for commercial  
CC. entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC. or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; M59810; AAA45468.1; -  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; -  
DR InterPro; IPR001205; - RNA\_pol; 1.  
DR Pfam; PR00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PR00910; RNA\_helicase; 1.  
DR Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 24 245

FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

Query Match 92.1%; Score 93; DB 1; Length 2226;  
Best Local Similarity 90.08; Pred. No. 1.6e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEHEHMKFS 20  
|||||  
Db 834 FSOAKISLFYTEHEHMKFS 853

RESULT 6  
POLG\_HPAV4  
ID POLG\_HPAV4 STANDARD; PRT; 2226 AA.  
AC P26581;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 43C).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M., Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination.";  
RT J. Virol. 65:2056-2065(1991).  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS. EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M59809; AAA45469.1; -  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; -  
DR InterPro; IPR001205; -  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Core protein; Transferase; RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
FT CHAIN 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B9BF75 CRC64;

Query Match 92.1%; Score 93; DB 1; Length 2226;  
Best Local Similarity 90.08; Pred. No. 1.6e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FSOAKISLFYTEHEHMKFS 20  
|||||  
Db 834 FSOAKISLFYTEHEHMKFS 853

RESULT 7  
POLG\_HPAV8  
ID POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 18f).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12096;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M., Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination.";  
RT J. Virol. 65:2056-2065(1991).  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS. EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M59808; AAA45467.1; -  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; -  
DR InterPro; IPR001205; -  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
KW Polyprotein; Coat protein; Core protein; Core protein; Transferase; RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
FT CHAIN 2226 AA; 251292 MW; 24964A63396C8D6B CRC64;

```

Query Match          92.1%; Score 93; DB 1; Length 2226;
Best Local Similarity 90.0%; Pred. No. 1.6e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEHMKFS 20
    ||||| ||||| ||||| ||||| |||||
DB 834 FSOAKISLFYTEHEHMKFS 853

RESULT 8
POLG_HPVS          STANDARD;          PRT: 2230 AA.
AC P14553;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Simian hepatitis A virus (strain AGM-27).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12102;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=91311420; PubMed=1649901;
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,
RA Purcell R.H.;
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome
RT structure and growth in cell culture with other HAV strains."
RL J. Gen. Virol. 72:1677-1683(1991).
[2]
SEQUENCE OF 1750-2164 FROM N.A.
RP MEDLINE=89232168; PubMed=2541023;
RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;
RT "Variations in genome fragments coding for RNA polymerase in human
RT and simian hepatitis A viruses."
RL FEBS Lett. 247:425-428(1989).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC
CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D00924; BAA00766.1;
DR EMBL: X15461; CAA33490.1;
DR PIR: A30470; GNNYSA.
DR PIR: S04885; S04885.
DR MEROPS: C03.005;
DR InterPro: IPR000605;
DR InterPro: IPR001205;
DR Pfam: PF00680; RNA_dep_RNA_pol.1.
DR Pfam: PF00910; RNA_helicase.1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 27 COAT PROTEIN VP4 (P1A).
FT CHAIN 28 249 COAT PROTEIN VP2 (P1B).
FT CHAIN 28 249 COAT PROTEIN VP3 (P1C).
FT CHAIN 250 495 COAT PROTEIN VP1 (P1D).
FT CHAIN 496 795 COAT PROTEIN P2A.
FT CHAIN 796 984 CORE PROTEIN P2B.
FT CHAIN 985 1091 CORE PROTEIN P2C.
FT CHAIN 1092 1426 CORE PROTEIN P2C.

Query Match          87.1%; Score 88; DB 1; Length 2230;
Best Local Similarity 85.0%; Pred. No. 1.1e-06;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEHEHMKFS 20
    ||||| ||||| ||||| ||||| |||||
DB 838 FSOAKISLFYTEHEHMKFS 857

RESULT 9
FDHA_METJA          STANDARD;          PRT: 378 AA.
ID Q60314;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE FORMATE DEHYDROGENASE ALPHA CHAIN (EC 1.2.1.2).
DE MJ0006.
GN Methanococcus jannaschii.
OS Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
[1]
SEQUENCE FROM N.A.
RP STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Adams M.D., Reich C.I.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Glodek A.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: FORMATE + NAD(+) = CO(2) + NADH
CC
CC -1- COFACTOR: MOLYBDENUM (MOLYBDOPTERIN), ZINC, FAD. MAY BIND A 4FE-4S
CC CLUSTER (BY SIMILARITY).
CC
CC -1- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS (BY SIMILARITY).
CC
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING
CC OXIDOREDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67459; AAB97987.1;
DR HSSP: P07658; 1FDI.
DR TIGR: M30006;
DR InterPro: IPR001467;
DR Pfam: PF00384; molybdopterin; 1.
DR PROSITE: PS00351; MOLYBDOPTERIN_PROK_1; 1.
DR PROSITE: PS00490; MOLYBDOPTERIN_PROK_2; 1.
DR PROSITE: PS00932; MOLYBDOPTERIN_PROK_3; FALSE NEG.
KW Hypothetical protein; Oxidoreductase; Zinc; Flavoprotein; Molybdenum;
KW FAD; NAD; Iron-sulfur; 4Fe-4S.
FT METAL 8 8 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 11 11 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 15 15 IRON-SULFUR (4FE-4S) (BY SIMILARITY).

```

```
FT METAL 43 43 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 378 AA; 42050 MW; E0C13060FF2609AA CRC64;

Query Match 46.5%; Score 47; DB 1; Length 378;
Best Local Similarity 47.1%; Pred. No. 1.4;
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 4 AKISLFYTEEHEIMKFS 20
DB 174 SKVTIFNTEKEIKLN 190

RESULT 10
YE28_CAEL STANDARD; PRT; 410 AA.
AC Q20585;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 47.6 KDA PROTEIN F49C12.8 IN CHROMOSOME IV.
GN F49C12.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OA NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gardner A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE FUS6 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z68227; CAA92512.1;
DR WormPep; F49C12.8; CE03368.
DR InterPro; IPR000717;
DR Pfam; PF01399; PCI: 1.
KW Hypothetical protein.
SQ SEQUENCE 410 AA; 47583 MW; F37390A3250109EE CRC64;

Query Match 42.6%; Score 43; DB 1; Length 410;
Best Local Similarity 36.8%; Pred. No. 7.3;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEEHEIMKF 19
DB 156 FAMIRVGLFFLDHLLNKF 174

RESULT 11
YC06_KLEPN STANDARD; PRT; 722 AA.
AC Q48452;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 80.4 KDA PROTEIN IN CPS REGION (ORF6).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OA NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CHEDI;

Query Match 41.1%; Score 41.5; DB 1; Length 288;
Best Local Similarity 41.7%; Pred. No. 8.8;
Matches 10; Conservative 5; Mismatches 4; Indels 5; Gaps 2;

QY 1 FSOAK--ISLFYTEE---HEIMKF 19
```

```
RX MEDLINE=95204345; PubMed=7896702;
RA Arakawa Y., Wacharotayankun R., Nagatsuka T., Ito H., Kato N.,
RT Ohta M.;
RT "Genomic organization of the Klebsiella pneumoniae cps region
RT responsible for serotype K2 capsular polysaccharide synthesis in the
RT virulent strain Chedi";
RL J. Bacteriol. 177:1788-1796(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D21242; BAA04777.1;
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 722 AA; 80400 MW; 3CAD6910AE81C3D7 CRC64;

Query Match 42.6%; Score 43; DB 1; Length 722;
Best Local Similarity 44.4%; Pred. No. 14;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEEHEIMK 18
DB 326 PRESEISQLYTKHEPTYK 343

RESULT 12
NIF3_YEAST STANDARD; PRT; 288 AA.
AC P53081;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NG1-INTERACTING FACTOR 3.
GN NIF3 OR YGL221C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OA NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII.";
RL Yeast 13:1077-1090(1997).
CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF3) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z72743; CAA96937.1;
DR SGD; S0003189; NIF3.
DR InterPro; IPR002678;
DR Pfam; PF01784; DUF34; 1.
SQ SEQUENCE 288 AA; 31888 MW; BGAB6E48AAAF4776A CRC64;

Query Match 41.1%; Score 41.5; DB 1; Length 288;
Best Local Similarity 41.7%; Pred. No. 8.8;
Matches 10; Conservative 5; Mismatches 4; Indels 5; Gaps 2;

QY 1 FSOAK--ISLFYTEE---HEIMKF 19
```

DB 217 FQKEDVDLYTGMESHVEVLK 240

RESULT 13

GCSPYRAB STANDARD; PRT: 138 AA.

AC Q9V0G1.1

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE PROBABLE GLYCINE CLEAVAGE SYSTEM H PROTEIN.

GN GCVH OR PAB0559.

OS Pyrococcus abyssi.

OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI\_TaxID=29292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORSAY;

RA Hellig R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYZES THE DEGRADATION OF GLYCINE. THE H PROTEIN SHUTTLES THE METHYLAMINE GROUP OF GLYCINE FROM THE P PROTEIN TO THE T PROTEIN (BY SIMILARITY).

CC -!- COPROFATOR: THE H CHAIN CONTAINS A COVALENTLY-BOUND LIPOYL COPROFATOR (BY SIMILARITY).

CC -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS: P, T, L, AND H (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE GCVH FAMILY.

CC -!- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: AJ248285; CAB49742.1; ALT\_INIT.

DR HSP; P16048; 1HTP.

DR InterPro: IPR002930; -.

DR InterPro: IPR003016; -.

DR Pfam: PF01597; GCV\_H; 1.

DR PROSITE: PS00189; LIPOYL; 1.

KW LIPOYL.

FT BINDING 74 74 LIPOYL (BY SIMILARITY).

SQ SEQUENCE 138 AA; 15545 MW; A015FC4B6F9A5792 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 138;

Best Local Similarity 53.8%; Pred. No. 4.7;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 QAKISLFYTEHE 15

DB 12 KVKEGLYTKE 24

RESULT 14

META\_THEME STANDARD; PRT: 304 AA.

AC Q9W2Y3.

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE HOMOSERINE O-SUCCINYLTRANSFERASE (EC 2.3.1.46) (HOMOSERINE O-TRANSSUCCINYLAASE) (HTS).

GN META OR TM0881.

OS Thermotoga maritima.

OC Bacteria; Thermotogales; Thermotoga.

OX NCBI\_TaxID=2336;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;

RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;

RT "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";

RL Nature 393:323-329(1999).

CC -!- CATALYTIC ACTIVITY: SUCCINYL-COA + L-HOMOSERINE = COA + O-SUCCINYL-L-HOMOSERINE.

CC -!- PATHWAY: BIOSYNTHESIS OF METHIONINE; HTS VARIANT; FIRST STEP.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE HTS FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: AE001753; AAD35962.1; -.

DR TIGR: TM0881; -.

KW Methionine biosynthesis; Transferase; Acyltransferase.

FT ACT\_SITE 142 142 POTENTIAL.

SQ SEQUENCE 304 AA; 35759 MW; 3ED8226A8F8044E CRC64;

Query Match 40.6%; Score 41; DB 1; Length 304;

Best Local Similarity 36.0%; Pred. No. 11;

Matches 9; Conservative 4; Mismatches 4; Indels 8; Gaps 1;

OY 3 QAKISLFYTEHE-----IMKF 19

DB 64 QVNVTLTETHKPKHTPIELK 88

RESULT 15

G6PLACICA STANDARD; PRT: 557 AA.

AC Q59088;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE GLUCOSE-6-PHOSPHATE ISOMERASE (GPI) (EC 5.3.1.9) (PHOSPHOGLUCOSE ISOMERASE) (PGI) (PHOSPHOHEXOSE ISOMERASE) (PHI).

GN PGI.

OS Acinetobacter calcoaceticus.

OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

OX Acinetobacter.

OX NCBI\_TaxID=471;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BD4;

RA Stark M., Kaplan N., Ron E.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: GLUCOSE 6-PHOSPHATE = FRUCTOSE 6-PHOSPHATE.

CC -!- PATHWAY: INVOLVED IN GLYCOLYSIS AND IN GLUCONEOGENESIS.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE GPI FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----

DR EMBL; X89900; CAA61993.1; -  
DR InterPro; IPR001672; -  
DR Pfam; PF00342; PGI; 1.  
DR PRINTS; PR00662; G6PIISOMERASE.  
DR PROSITE; PS00765; P\_GLUCOSE\_ISOMERASE\_1; 1.  
DR PROSITE; PS00174; P\_GLUCOSE\_ISOMERASE\_2; 1.  
KW Gluconeogenesis; Glycolysis; Isomerase.  
SQ SEQUENCE 557 AA; 63280 MW; D74AF214B139E4DC CRC64;

Query Match 40.6%; Score 41; DB 1; Length 557;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 10; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

Qy 1 FSQAKISLFYTEEHIMKFS 20  
||| ||: |||: | |:  
Db 87 FSQNKIN--YTEQREAHWA 104

Search completed: April 24, 2001, 16:34:03  
Job time: 559 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 24, 2001, 16:33:10 ; Search time 113.97 Seconds  
(without alignments)  
20.568 Million cell updates/sec

Title: US-09-171-432A-43  
Perfect score: 101  
Sequence: 1 FSQAKISLFYTEHEIMKFS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mhc:\*
- 8: sp-organelle:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-unclassified:\*
- 13: sp-vertebrate:\*
- 14: sp-virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	101	100.0	112	14 Q86534	Q86534 hepatitis a
2	101	100.0	1124	14 Q84780	Q84780 hepatitis a
3	101	100.0	1161	14 Q05794	Q05794 hepatitis a
4	101	100.0	2216	14 Q9WMA2	Q9WMA2 hepatitis a
5	101	100.0	2218	14 Q67824	Q67824 hepatitis a
6	101	100.0	2218	14 Q67817	Q67817 hepatitis a
7	101	100.0	2227	14 Q67825	Q67825 hepatitis a
8	101	100.0	2227	14 Q67826	Q67826 hepatitis a
9	101	100.0	2227	14 Q9WMA4	Q9WMA4 hepatitis a
10	101	100.0	2227	14 Q9WMA3	Q9WMA3 hepatitis a
11	101	100.0	2227	14 Q9WMA1	Q9WMA1 hepatitis a
12	101	100.0	2227	14 Q9WMA0	Q9WMA0 hepatitis a
13	101	100.0	2227	14 Q9WMA9	Q9WMA9 hepatitis a
14	98	97.0	184	14 Q87092	Q87092 simian hepa
15	98	97.0	2227	14 Q9IFH5	Q9IFH5 hepatitis a
16	49	48.5	116	14 Q71975	Q71975 hepatitis a
17	49	48.5	116	14 Q71977	Q71977 hepatitis a
18	49	48.5	116	14 Q71978	Q71978 hepatitis a
19	49	48.5	116	14 Q9W7X8	Q9W7X8 hepatitis a

```

20 49 48.5 116 14 Q9W7X7 hepatitis a
21 49 48.5 116 14 Q9W7T4 hepatitis a
22 49 48.5 116 14 Q9W7T0 hepatitis a
23 49 48.5 116 14 Q9W7S7 hepatitis a
24 49 48.5 116 14 Q92941 hepatitis a
25 48 47.5 505 8 Q98636 hepatitis a
26 47 46.5 1587 10 Q9LQES hepatitis a
27 45 44.6 721 2 Q9X4B9 hepatitis a
28 44 43.6 503 8 Q98638 hepatitis a
29 44 43.6 507 8 Q98639 hepatitis a
30 44 43.6 509 8 Q98637 hepatitis a
31 43 42.6 325 2 Q98KJ5 hepatitis a
32 43 42.6 510 8 Q9TKJ3 hepatitis a
33 42 41.6 115 14 Q71976 hepatitis a
34 42 41.6 187 5 Q17580 hepatitis a
35 42 41.6 306 8 Q35136 hepatitis a
36 42 41.6 1033 10 Q9LG34 hepatitis a
37 42 41.6 1265 5 Q9XUE3 hepatitis a
38 41.5 41.1 1327 5 Q9NFB4 hepatitis a
39 41 40.6 144 1 Q9VOG1 hepatitis a
40 41 40.6 209 3 Q08981 hepatitis a
41 41 40.6 238 5 Q9VMP5 hepatitis a
42 41 40.6 304 2 Q9WY3 hepatitis a
43 41 40.6 483 2 Q9K6L8 hepatitis a
44 41 40.6 820 10 Q9LFE0 hepatitis a
45 41 40.6 1336 14 Q73550 hepatitis a

```

#### ALIGNMENTS

```

RESULT 1
Q86534 ID Q86534 PRELIMINARY: PRT: 112 AA.
AC Q86534;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE VP1 PROTEIN (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92348853; PubMed=1668326;
RA Fineschi N., Cavalieri F., Garelick H., Prugnola A., Pellegrini V.,
RA Zuckerman A.J.;
RT "Characterization of a hepatitis A virus strain suitable for vaccine
RT production.";
RL J. Hepatol. 13 Suppl. 4:146-151(1991).
DR EMBL; S44105; CAB28369.1; -
FT NON_TER 112 112
SQ SEQUENCE 112 AA; 13145 MW; 3C0CBBB4570D9A2C CRC64;

```

Query Match 100.0%; Score 101; DB 14; Length 112;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 FSQAKISLFYTEHEIMKFS 20
Db 75 FSQAKISLFYTEHEIMKFS 94

```

```

RESULT 2
Q84780 ID Q84780 PRELIMINARY: PRT: 1124 AA.
AC Q84780;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS)

```

```
DE (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Ovchinnikov Y.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,
RA Rokhlina T.O., Chizhikov V.E., Petrov N.A., Prikhod'ko G.G.,
RA Blinov V.M., Vasilenko S.K., Sandakchiev L.S., Kusov Y.Y.,
RA Grabko V.I., Fleer G.P., Balyan M.S., Drozdov S.G.;
RL Dokl. Biochem. 285:379-383(1986).
DR EMBL; X04200; CAA27797.1; -.
DR EMBL; All312; CAA00953.1; -.
KW Nonstructural protein.
FT NON_TER 1
FT NON_TER 1124
SQ SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

Query Match 100.0%; Score 101; DB 14; Length 1124;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FSQAKISLFYTEEHEIMKFS 20
| | | | | | | | | | | | | | | | | |
Db 791 FSQAKISLFYTEEHEIMKFS 810

RESULT 3
ID Q05794 PRELIMINARY; PRT; 1161 AA.
AC Q05794; Q67800; Q67801; Q67802; Q67803; Q67804; Q67805; Q67806;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE GENOME POLYPEPTIDE (COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)) (FRAGMENT).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilenko S.K.,
RA Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,
RA Balayan M.S.;
RL Mol. Gen. Microbiol. Virol. 6:129-133(1987).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS.
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL; X15464; CAA33492.1; -.
DR INTERPRO; IPR000408; -.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein; Coat protein; RNA-directed RNA polymerase;
KW Hydrolase; Thiol protease.
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).
FT CHAIN 795 900 CORE PROTEIN P2A.
FT CHAIN 901 1087 CORE PROTEIN P2B.
FT CHAIN 1088 >1161 CORE PROTEIN P2C.
FT NON_TER 1161 1161
SQ SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 1161;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FSQAKISLFYTEEHEIMKFS 20
```

```
Db 828 FSQAKISLFYTEEHEIMKFS 847
| | | | | | | | | | | | | | | | | |
RESULT 4
ID Q9WMA2 PRELIMINARY; PRT; 2216 AA.
AC Q9WMA2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RA Fujiwara K.;
RT "Hepatitis A virus";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB020566; BAA35104.1; -.
DR INTERPRO; IPR000408; -.
DR INTERPRO; IPR000605; -.
DR INTERPRO; IPR001205; -.
DR INTERPRO; IPR001643; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSN.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2216 AA; 250209 MW; 1A9D93FEC21FBE82 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2216;
Best Local Similarity 100.0%; Pred. No. 6.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FSQAKISLFYTEEHEIMKFS 20
| | | | | | | | | | | | | | | | | |
Db 834 FSQAKISLFYTEEHEIMKFS 853

RESULT 5
ID Q67824 PRELIMINARY; PRT; 2218 AA.
AC Q67824;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE GBM/FRHK RNA.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GBM;
RX MEDLINE=94076453; PubMed=8254770;
RA Graff J., Normann A., Feinstein S.M., Flehmig B.;
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison
RT with two cell culture-adapted variants.";
RL J. Virol. 68:548-554(1994).
DR EMBL; X75214; CAA53024.1; -.
DR INTERPRO; IPR000408; -.
DR INTERPRO; IPR000605; -.
DR INTERPRO; IPR001205; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
FT CHAIN 1 785 P1 STRUCTURAL PROTEINS.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEINS.
```



FT CHAIN 1417 2218 P3 NONSTRUCTURAL PROTEINS.  
SQ SEQUENCE 2218 AA; 250502 MW; CA72DF0922104C0E CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2218;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSQAKISLFYTEEHEIMKFS 20  
|||||  
Db 828 FSQAKISLFYTEEHEIMKFS 847

RESULT 6

O67817 PRELIMINARY; PRT; 2218 AA.

AC Q67817;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE COMPLETE GENOME.

OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.  
OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-F.G.;

RX MEDLINE=95381623; PubMed=7653108;

RA Beneduce F., Pisani G., Divizia M., Pana A., Morace G.;

RT "Complete nucleotide sequence of a cytopathic hepatitis A virus strain isolated in Italy.";

RL Virus Res. 36:299-309(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-F.G.;

RA Morace G.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

DR EMBL; X83302; CAA58281.1; -;

DR INTERPRO; IPR000408; -;

DR INTERPRO; IPR000605; -;

DR INTERPRO; IPR001205; -;

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PFAM; PF00910; RNA\_helicase; 1.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

SQ SEQUENCE 2218 AA; 250476 MW; 813821D3E4E533CA CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2218;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSQAKISLFYTEEHEIMKFS 20  
|||||  
Db 828 FSQAKISLFYTEEHEIMKFS 847

RESULT 7

O67825 PRELIMINARY; PRT; 2227 AA.

AC Q67825;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE GBM/WT RNA.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-GBM;

RX MEDLINE=94076453; PubMed=8254770;

RA Graff J., Normann A., Feinstone S.M., Flehmig B.;

RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison with two cell culture-adapted variants.";

RL J. Virol. 68:548-554(1994).

DR EMBL; X75215; CAA53025.1; -;

DR INTERPRO; IPR000408; -;

DR INTERPRO; IPR000605; -;

DR INTERPRO; IPR001205; -;

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PFAM; PF00910; RNA\_helicase; 1.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.

FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.

FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.

SQ SEQUENCE 2227 AA; 251563 MW; 4C4D79D352F936B4 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSQAKISLFYTEEHEIMKFS 20  
|||||

Db 834 FSQAKISLFYTEEHEIMKFS 853

RESULT 8

O67826

ID Q67826 PRELIMINARY; PRT; 2227 AA.

AC Q67826;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)

DE GBM/HES RNA.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-GBM;

RX MEDLINE=94076453; PubMed=8254770;

RA Graff J., Normann A., Feinstone S.M., Flehmig B.;

RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison with two cell culture-adapted variants.";

RL J. Virol. 68:548-554(1994).

DR EMBL; X75216; CAA53026.1; -;

DR INTERPRO; IPR000408; -;

DR INTERPRO; IPR000605; -;

DR INTERPRO; IPR001205; -;

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PFAM; PF00910; RNA\_helicase; 1.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.

FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.

FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.

SQ SEQUENCE 2227 AA; 251496 MW; 48CB7C962319457 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FSQAKISLFYTEEHEIMKFS 20  
|||||

Db 834 FSQAKISLFYTEEHEIMKFS 853

RESULT 9

Q9WMA4

ID Q9WMA4 PRELIMINARY; PRT; 2227 AA.

AC Q9WMA4;

DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-AH1;

RA Fujiwara K.;

RT "Hepatitis A virus.";

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB020564; BAA35102.1; -

DR INTERPRO; IPR000408; -

DR INTERPRO; IPR000605; -

DR INTERPRO; IPR001205; -

DR INTERPRO; IPR001643; -

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PRINTS; PR00910; RNA\_helicase; 1.

DR PRINTS; PR00918; CALICVIRUSNS.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

KW Polyprotein.

SQ SEQUENCE 2227 AA; 251304 MW; 0DEF6D2AEC29C0CE CRC64;

Query Match

Best Local Similarity 100.0%; Score 101; DB 14; Length 2227;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FSOAKISLFYTEEHEIMKFS 20

|||||

Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 10

Q9WMA3

ID Q9WMA3 PRELIMINARY; PRT; 2227 AA.

AC Q9WMA3;

DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-AH2;

RA Fujiwara K.;

RT "Hepatitis A virus.";

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB020565; BAA35103.1; -

DR INTERPRO; IPR000408; -

DR INTERPRO; IPR000605; -

DR INTERPRO; IPR001205; -

DR INTERPRO; IPR001643; -

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PRINTS; PR00910; RNA\_helicase; 1.

DR PRINTS; PR00918; CALICVIRUSNS.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

KW Polyprotein.

SQ SEQUENCE 2227 AA; 251440 MW; E04F846CEC7F50FD CRC64;

Query Match

Best Local Similarity 100.0%; Score 101; DB 14; Length 2227;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FSOAKISLFYTEEHEIMKFS 20

|||||

Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 11

Q9WMA1

ID Q9WMA1 PRELIMINARY; PRT; 2227 AA.

AC Q9WMA1;

DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-FH1;

RA Fujiwara K.;

RT "Hepatitis A virus.";

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB020567; BAA35105.1; -

DR INTERPRO; IPR000408; -

DR INTERPRO; IPR000605; -

DR INTERPRO; IPR001205; -

DR INTERPRO; IPR001643; -

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PRINTS; PR00910; RNA\_helicase; 1.

DR PRINTS; PR00918; CALICVIRUSNS.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

KW Polyprotein.

SQ SEQUENCE 2227 AA; 251415 MW; F92C8E2323FC5621 CRC64;

Query Match

Best Local Similarity 100.0%; Score 101; DB 14; Length 2227;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FSOAKISLFYTEEHEIMKFS 20

|||||

Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 12

Q9WMA0

ID Q9WMA0 PRELIMINARY; PRT; 2227 AA.

AC Q9WMA0;

DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-FH2;

RA Fujiwara K.;

RT "Hepatitis A virus.";

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB020568; BAA35106.1; -

DR INTERPRO; IPR000408; -

DR INTERPRO; IPR000605; -

DR INTERPRO; IPR001205; -

DR INTERPRO; IPR001643; -

DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.

DR PRINTS; PR00910; RNA\_helicase; 1.

DR PRINTS; PR00918; CALICVIRUSNS.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

KW Polyprotein.

SQ SEQUENCE 2227 AA; 251118 MW; 53E86B4432127E9B CRC64;

```
Query Match      100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEEHEIMKFS 20
   |||||
Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 13
Q9WM99          PRELIMINARY;      PRT; 2227 AA.
AC Q9WM99;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FH3;
RA Fujiwara K.;
RT "Hepatitis A virus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB020569; BAA35107.1; -.
DR INTERPRO; IPR000408; -.
DR INTERPRO; IPR000605; -.
DR INTERPRO; IPR001205; -.
DR INTERPRO; IPR001643; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSN.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251387 MW; 817640187672A23D CRC64;

Query Match      100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 6.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEEHEIMKFS 20
   |||||
Db 834 FSOAKISLFYTEEHEIMKFS 853

RESULT 14
Q87092          PRELIMINARY;      PRT; 184 AA.
AC Q87092;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE SEGMENT 2A-ENCODED PROTEIN (FRAGMENT).
OS Simian hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90012354; PubMed=2552172;
RA Brown E.A.; Jansen R.W.; Lemon S.M.;
RT "Characterization of a simian hepatitis A virus (HAV): antigenic and
   genetic comparison with human HAV.";
RL J. Virol. 63:4932-4937(1989).
DR EMBL; M34085; AAA47483.1; -.
DR NON_TER 1
FT NON_TER 184
SQ SEQUENCE 184 AA; 21322 MW; 9FA851FA964A88A8 CRC64;
```

```
Query Match      97.0%; Score 98; DB 14; Length 184;
Best Local Similarity 95.0%; Pred. No. 1.4e-08;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEEHEIMKFS 20
   |||||
Db 43 FSOAKISLFYTEEHEIMKFS 62

RESULT 15
Q9IFH5          PRELIMINARY;      PRT; 2227 AA.
ID Q9IFH5;
AC Q9IFH5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE POLYPROTEIN PRECURSOR.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HAF-203;
RA Baptista M.L.; Silva M.; de Lima M.A.; Yoshida C.F.; Gaspar A.M.;
RA Pires Lopes M.Q.; Galler R.;
RT "Nucleotide sequence of the HAF-203 hepatitis A virus strain isolated
   in Brazil and expression of the VP1 gene in a bacterial system.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF288396; AAF80114.1; -.
DR SEQUENCE 2227 AA; 251432 MW; 81913AECC8A04200 CRC64;

Query Match      97.0%; Score 98; DB 14; Length 2227;
Best Local Similarity 95.0%; Pred. No. 1.9e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSOAKISLFYTEEHEIMKFS 20
   |||||
Db 834 FSOAKISLFYTEEHEIMKFS 853
```

Search completed: April 24, 2001, 16:33:13  
Job time: 574 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:23:27 ; Search time 121.9 seconds  
(without alignments)  
9.379 Million cell updates/sec

Title: US-09-171-432a-44

Perfect score: 104

Sequence: 1 KVNPPHGMLEETAAANSKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0401.\*  
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	20	18 W42927	Immunogenic Hepati
2	104	100.0	2227	7 P60066	Sequence of viral
3	104	100.0	2227	11 R05697	Attenuated Hepatit
4	104	100.0	2227	18 W34074	Hepatitis A virus
5	104	100.0	2227	21 B18607	Amino acid sequenc
6	104	100.0	2227	21 B18608	Amino acid sequenc
7	104	100.0	2227	21 B18609	Amino acid sequenc
8	53	51.0	20	18 W42928	Immunogenic Hepati
9	45	43.3	206	21 G32341	Arabidopsis thalia
10	45	43.3	242	21 G10097	Arabidopsis thalia
11	45	43.3	242	21 G32340	Arabidopsis thalia

12	45	43.3	383	21	G10096	Arabidopsis thalia
13	45	43.3	383	21	G32339	Arabidopsis thalia
14	45	43.3	399	21	G10095	Arabidopsis thalia
15	44	42.3	432	20	Y9167	Human succinyl-Coe
16	44	42.3	432	21	Y76886	Human SCSH-2 prote
17	43.5	41.8	267	21	G37721	Arabidopsis thalia
18	43.5	41.8	267	21	G37736	Arabidopsis thalia
19	43.5	41.8	269	21	G24210	Arabidopsis thalia
20	43.5	41.8	426	21	G37720	Arabidopsis thalia
21	43.5	41.8	426	21	G37735	Arabidopsis thalia
22	43.5	41.8	428	21	G24209	Arabidopsis thalia
23	43.5	41.8	438	21	G37734	Arabidopsis thalia
24	43.5	41.8	440	21	G24208	Arabidopsis thalia
25	43.5	41.8	442	21	G37719	Arabidopsis thalia
26	43	41.3	98	19	W29645	Human secreted pro
27	43	41.3	98	20	X39330	Protein product of
28	43	41.3	98	21	B44333	Human PRO725 prote
29	43	41.3	98	22	B49893	Human secreted pro
30	43	41.3	99	20	Y02719	Human secreted pro
31	43	41.3	110	20	X36052	Extended human sec
32	43	41.3	124	18	W20410	H. pylori cytoplas
33	43	41.3	124	18	W24654	H. pylori cytoplas
34	43	41.3	151	20	X37664	Protein which is s
35	43	41.3	443	18	W20743	H. pylori cytoplas
36	41	39.4	226	21	G22472	Arabidopsis thalia
37	41	39.4	239	21	G22471	Arabidopsis thalia
38	41	39.4	244	21	G22470	Arabidopsis thalia
39	41	39.4	263	21	B56534	Human prostate can
40	41	39.4	373	21	G36606	Arabidopsis thalia
41	41	39.4	396	21	G36605	Arabidopsis thalia
42	41	39.4	470	21	G36604	Arabidopsis thalia
43	40	38.5	194	21	G24792	Arabidopsis thalia
44	40	38.5	589	20	Y05829	Yeast sphingosine-
45	40	38.5	835	21	G31412	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1  
W42927  
ID W42927 standard; peptide; 20 AA.  
AC W42927;  
XX  
XX W42927;  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1327.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW Immune response; antibody.  
XX  
XX Synthetic.  
OS Hepatitis A virus.  
XX  
PN WO9740147-A1.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US06891.  
XX  
PR 19-APR-1996; 96US-0015644.  
XX  
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.  
PA Fields HA, Khudyakov YE;  
PI Amino acid sequenc  
XX  
DR Amino acid sequenc  
XX  
XX Amino acid sequenc  
PT Immunogenic Hepati  
PT Arabidopsis thalia  
PT Arabidopsis thalia  
XX

Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an immune response to HAV in a mammal or to detect the presence of antibodies against HAV in a mammal

PS Claim 18; Page 112; 140pp; English.

XX Peptides W42922-30 are immunogenic peptides corresponding to immunogenic  
CC epitopes of the Hepatitis A virus (HAV). The peptides are substantially  
CC similar to a portion of the amino acid sequence of the P2A protein of HAV  
CC corresponding to amino acids 792-980. The present peptide is derived  
CC from amino acids 922-941, and has a reactivity of 31.3% with acute sera.  
CC Compositions containing the peptides can be used to induce an immune  
CC response to HAV in a mammal. The peptides can also be used to detect the  
CC presence of antibodies against HAV in mammalian serum. The peptides can  
CC also be used to make an antibody against HAV by administering the peptide  
CC to a mammal.

XX Sequence 20 AA;

Query Match 100.0%; Score 104; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.2e-11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDLEIEIAANSKD 20

Db 1 kvnfphgmldleelaanskd 20

RESULT 2

P60066  
ID P60066 standard; Protein; 2227 AA.

XX AC P60066;

XX DT 26-JUN-1991 (first entry)

XX DE Sequence of viral L434 polypeptide encoded by the complete  
DE nucleotide sequence of the HAV genome.

XX KW Diagnosis; vaccine; passive immunotherapy.

XX OS Hepatitis A virus.

XX FH Key Location/Qualifiers

FT 1..245 /label= P1.1A

FT 246..491 /label= 1B

FT 492..836 /label= 1C

FT 837..980 /label= 1C

FT 981..1076 /label= P2.2A

FT 1077..1422 /label= 2B

FT 1423..1484 /label= 2C

FT 1485..1507 /label= P3.3A

FT 1508..1678 /label= 3B

FT 1679..2227 /label= 3C

FT 1679..2227 /label= 3D

XX EP199480-A.

XX PD 29-OCT-1986.

XX PF 03-APR-1986; 86EP-0302465.

XX PR 03-APR-1985; 85US-0719329.

XX FA (CHIR-) CHIRON CORP.

XX PI Dina D, Potter SJ, Vannest GA, Caput D;

XX WPI: 1986-286213/44.  
DR N-PSDB; N60080.

XX Hepatitis A virus nucleotide sequence and polypeptide - and use  
PT in prodn. of vaccines and diagnostic probes

XX Claim 5; Fig 1; 18pp; English.

XX N60080 and oligonucleotide fragments are useful in detection of  
CC hepatitis A virus; transformed hosts may be used for expression of  
CC polypeptides and fragments useful in vaccines without risk of  
CC infection by the virus or in prodn. of particles which are capable  
CC of inducing immunocompetent B cells for passive immunotherapy. Pref.  
CC epitope is derived from AAs 445-657 or 792-848 of the HAV  
CC polypeptide sequence (P60066).

XX Sequence 2227 AA;

Query Match 100.0%; Score 104; DB 7; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 9.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDLEIEIAANSKD 20

Db 922 kvnfphgmldleelaanskd 941

RESULT 3

R05697  
ID R05697 standard; protein; 2227 AA.

XX AC R05697;

XX DT 15-AUG-1990 (first entry)

XX DE Attenuated hepatitis A virus.

XX KW Hepatitis A virus; vaccine; attenuated.

XX OS Hepatitis A virus, strain HM-175.

XX FH Key Location/Qualifiers

FT 1..23 /label=VP4 = 1A

FT 24..245 /label=VP2 = 1B

FT 246..491 /label=VP3 = 1C

FT 492..791 /label=VP1 = 1D

FT 792..980 /label=2A

FT 981..1087 /label=2B

FT 1088..1422 /label=2C

FT 1423..1496 /label=3A

FT 1497..1519 /label=3B = VPg

FT 1520..1738 /label=3C

FT 1739..2227 /label=3D

XX US4894228-A.

XX PD 16-JAN-1990.

XX PF 12-JUL-1988; 88US-0217824.

```

PR 12-JUL-1988: 88US-0217824.
XX 12-JUL-1988: 88US-0652967.
XX (USSH ) US DEPT HEALTH & HUMAN.
XX Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstone SM;
PI Daemer RJ, Gust ID;
XX DR N-PSDB; Q03512.
XX WPI; 1990-075557/10.
XX DR N-PSDB; Q03512.
XX Vaccine against hepatitis A virus infection - comprises novel
PT attenuated hepatitis A virus strain.
XX Claim 1; Fig 1; 18pp; English.
XX The attenuated HAV is useful for inducing protective immunity against
CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by
CC several nucleotide changes distributed throughout the genome, is
CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is
CC suitable for use as an HAV vaccine. It is noted that not all the changes
CC are necessary for attenuation and use as a vaccine.
XX Sequence 2227 AA;
SQ
Query Match 100.0%; Score 104; DB 11; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KVNPHGMLDLEIAANSKD 20
DB 922 kvnphgmldleelaanskd 941
|||||
RESULT 4
W34074
ID W34074 standard; Protein: 2227 AA.
XX W34074;
XX
XX 27-APR-1998 (first entry)
XX Hepatitis A virus HM-175 protein sequence.
XX DE Hepatitis A virus HM-175 protein sequence.
XX KW HAV; attenuation: 2C protein; 2C chimeric hepatitis A virus;
XX infection; vaccine.
XX OS Hepatitis A virus HM-175.
XX
XX Key Location/Qualifiers
FH Protein 1..23
FT /label= VP4
FT Protein 24..245
FT /label= VP2
FT Protein 246..491
FT /label= VP3
FT Protein 492..791
FT /label= VP1
FT Protein 792..980
FT /label= 2A
FT Protein 981..1087
FT /label= 2B
FT Protein 1088..1422
FT /label= 2C
FT Protein 1423..1496
FT /label= 3A
FT Protein 1497..1519
FT /label= 3B
FT Protein 1520..1738
FT /label= 3C
FT Protein 1739..2227
FT /label= 3D

```

```

XX WO9740166-A2.
XX 30-OCT-1997.
XX 18-APR-1997; 97WO-US06506.
XX 19-APR-1996; 96US-0015642.
XX (USSH ) US SEC DEPT HEALTH.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX Emerson SU, Purcell RH, Raychaudhuri G;
XX WPI; 1997-535850/49.
XX N-PSDB; T93023.
XX Human attenuated HAV genome containing simian HAV 2C gene - useful
PT as vaccines against HAV infection
XX Disclosure; Fig 13A-D; 66pp; English.
XX This protein sequence is encoded by the human hepatitis A virus
CC (HAV) HM-175 wild-type genome (see T93023). Attenuated strain
CC HAV/7 is obtained by passage of HM-175 in African Green Monkey
CC kidney cells. A claimed DNA construct (1) comprises a genome of
CC HAV, where the genome is a human attenuated HAV genome in which a
CC region of the 2C gene has been replaced by a corresponding region
CC from a 2C gene of a simian AGM-27 HAV genome (see T93024). The
CC region of the 2C gene from AGM-27 contained in the construct
CC preferably encodes amino acids 120-328 of the 2C protein, amino
CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA
CC transcript of (1); (2) a cell transfected with (1) or the RNA
CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the
CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or
CC its RNA transcript, can be used as a vaccine for preventing HAV in
CC a mammal. (1) or the RNA transcript can also be used to stimulate
CC the production of protective antibodies in the mammal.
XX Sequence 2227 AA;
SQ
Query Match 100.0%; Score 104; DB 18; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KVNPHGMLDLEIAANSKD 20
DB 922 kvnphgmldleelaanskd 941
|||||
RESULT 5
B18607
ID B18607 standard; Protein: 2227 AA.
XX B18607;
XX 15-JAN-2001 (first entry)
XX Amino acid sequence of wild type Hepatitis A virus strain HM-175.
XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.
XX Hepatitis A virus.
XX US6113912-A.
XX 05-SEP-2000.
XX 07-JUN-1995; 95US-0475886.
XX 18-SEP-1992; 92US-0947338.
XX 17-SEP-1993; 93WO-US08610.
PR

```

PR 10-MAR-1995; 95US-0397232.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX WPI; 2000-586464/55.  
DR N-PSDB; A75476.  
XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type -  
XX Disclosure; Fig 6A-K; 72pp; English.  
XX The present sequence is derived from a wild type hepatitis A virus  
CC (HAV) strain HM-174. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.  
XX Sequence 2227 AA;  
SQ

Query Match 100.0%; Score 104; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 9.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDLEETAAANSKD 20  
Db 922 kvnfphgmldeelaanskd 941

RESULT 6  
BI8608  
ID BI8608 standard; Protein; 2227 AA.  
XX  
AC BI8608;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.  
XX  
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
KW P-35 virus.  
XX  
OS Hepatitis A virus.  
XX  
PN US6113912-A.  
XX  
PD 05-SEP-2000.  
XX  
PF 07-JUN-1995; 95US-0475886.  
XX  
PR 18-SEP-1992; 92US-0947338.  
PR 17-SEP-1993; 93WO-US08610.  
PR 10-MAR-1995; 95US-0397232.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX  
DR WPI; 2000-586464/55.  
DR N-PSDB; A75477.  
XX  
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type -  
XX Disclosure; Columns 67-78; 72pp; English.  
XX  
CC The present sequence is derived from passage 35 of a wild type

CC hepatitis A virus (HAV) strain HM-174. The resulting virus is  
CC designated P-35 virus. The sequence is modified to produce HAV which  
CC are adapted to growth in the human fibroblast-like cell line MRC-5.  
CC The HAV is able to propagate in MRC-5 cells and retain appropriate  
CC attenuation. It is useful as a live vaccine for prophylaxis of  
CC hepatitis A in humans and other primates.  
XX Sequence 2227 AA;  
SQ

Query Match 100.0%; Score 104; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 9.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDLEETAAANSKD 20  
Db 922 kvnfphgmldeelaanskd 941

RESULT 7  
BI8609  
ID BI8609 standard; Protein; 2227 AA.  
XX  
AC BI8609;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE Amino acid sequence of live attenuated Hepatitis A virus 4380.  
XX  
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
KW HAV 4380.  
XX  
OS Hepatitis A virus.  
XX  
PN US6113912-A.  
XX  
PD 05-SEP-2000.  
XX  
PF 07-JUN-1995; 95US-0475886.  
XX  
PR 18-SEP-1992; 92US-0947338.  
PR 17-SEP-1993; 93WO-US08610.  
PR 10-MAR-1995; 95US-0397232.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX  
DR WPI; 2000-586464/55.  
DR N-PSDB; A75478.  
XX  
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type -  
XX Disclosure; Columns 93-104; 72pp; English.  
XX  
CC The present sequence is derived from a live attenuated hepatitis A  
CC virus (HAV) of the invention, designated HAV 4380. The sequence is  
CC produced by modifying wild type HAV strain HM-174. The HAV of the  
CC invention are adapted to growth in the human fibroblast-like cell  
CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain  
CC appropriate attenuation. It is useful as a live vaccine for prophylaxis  
CC of hepatitis A in humans and other primates.  
XX Sequence 2227 AA;  
SQ

Query Match 100.0%; Score 104; DB 21; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 9.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDLEETAAANSKD 20



```
Db      922 knvphgmldleeaanskd 941
|||||
RESULT      8
W42928
ID      W42928 standard; peptide; 20 AA.
XX
AC      W42928;
XX
DT      28-APR-1998 (first entry)
XX
DE      Immunogenic Hepatitis A virus peptide YK-1328.
XX
KW      Immunogenic peptide; immunogenic epitope; P2A protein;
KW      Immune response; antibody.
XX
OS      Synthetic.
OS      Hepatitis A virus.
XX
PN      WO9740147-A1.
XX
PD      30-OCT-1997.
XX
PF      18-APR-1997; 97WO-US06891.
XX
PR      19-APR-1996; 96US-0015644.
XX
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Fields HA, Khudyakov YE;
XX
DR      WPI; 1997-535831/49.
XX
PT      Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an
PT      immune response to HAV in a mammal or to detect the presence of
PT      antibodies against HAV in a mammal
XX
PS      Claim 18; Page 112; 140pp; English.
XX
CC      Peptides W42922-30 are immunogenic peptides corresponding to immunogenic
CC      epitopes of the Hepatitis A virus (HAV). The peptides are substantially
CC      similar to a portion of the amino acid sequence of the P2A protein of HAV
CC      corresponding to amino acids 792-980. The present peptide is derived
CC      from amino acids 931-950, and has a reactivity of 12.5% with acute sera.
CC      Compositions containing the peptides can be used to induce an immune
CC      response to HAV in a mammal. The peptides can also be used to detect the
CC      presence of antibodies against HAV in mammalian serum. The peptides can
CC      also be used to make an antibody against HAV by administering the peptide
CC      to a mammal.
XX
SQ      Sequence 20 AA;

Query Match      51.0%; Score 53; DB 18; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 DLEEIAANSKD 20
|||||
Db      1 dleeiaanskd 11
|||||

RESULT      9
G32341
ID      G32341 standard; Protein; 206 AA.
XX
AC      G32341;
XX
DT      17-OCT-2000 (first entry)
XX
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 38996.
XX
```

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EPI033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

99US-0121825.

99US-0123180.

99US-0123548.

99US-0125788.

99US-0126264.

99US-0126785.

99US-0127462.

99US-0128234.

99US-0128714.

99US-0129845.

99US-0130077.

99US-0130449.

99US-0130510.

99US-0130891.

99US-0131449.

99US-0132048.

99US-0132407.

99US-0132484.

99US-0132485.

99US-0132486.

99US-0132487.

99US-0132863.

99US-0132863.

99US-0134256.

99US-0134218.

99US-0134219.

99US-0134221.

99US-0134370.

99US-0134370.

99US-0134768.

99US-0134941.

99US-0135124.

99US-0135353.

99US-0135629.

99US-0136021.

99US-0136392.

99US-0136782.

99US-0137222.

99US-0137528.

99US-0137502.

99US-0137724.

99US-0138094.

99US-0138540.

99US-0138847.

99US-0139119.

99US-0139452.

99US-0139453.

99US-0139492.

99US-0139454.

99US-0139455.

99US-0139456.

99US-0139457.

99US-0139458.

99US-0139459.

99US-0139460.

99US-0139461.

99US-0139462.

99US-0139463.

99US-0139750.

99US-0139763.

99US-0139817.

99US-0139899.

99US-0140353.

```

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.3%; Score 45; DB 21; Length 206;
Best Local Similarity 44.4%; Pred. No. 6.2;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 NPHGMLDLREIAANSKD 20
Db 32 dypvgllidiekilppgk 49

RESULT 10
GI0097
ID GI0097 standard; Protein; 242 AA.
XX
AC GI0097;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8283.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
```

XX	06-SEP-2000.	PR	02-JUL-1999;	99US-0142055.
PD		PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
PF	25-FEB-2000; 2000EP-0301439.	PR	09-JUL-1999;	99US-0142920.
XX		PR	12-JUL-1999;	99US-0142977.
PR		PR	13-JUL-1999;	99US-0143542.
PR		PR	14-JUL-1999;	99US-0143624.
PR		PR	15-JUL-1999;	99US-0144005.
PR		PR	16-JUL-1999;	99US-0144085.
PR		PR	16-JUL-1999;	99US-0144086.
PR		PR	19-JUL-1999;	99US-0144325.
PR		PR	19-JUL-1999;	99US-0144331.
PR		PR	19-JUL-1999;	99US-0144332.
PR		PR	19-JUL-1999;	99US-0144333.
PR		PR	19-JUL-1999;	99US-0144334.
PR		PR	19-JUL-1999;	99US-0144335.
PR		PR	20-JUL-1999;	99US-0144352.
PR		PR	20-JUL-1999;	99US-0144632.
PR		PR	20-JUL-1999;	99US-0144632.
PR		PR	21-JUL-1999;	99US-0144884.
PR		PR	21-JUL-1999;	99US-0144814.
PR		PR	21-JUL-1999;	99US-0145086.
PR		PR	21-JUL-1999;	99US-0145088.
PR		PR	22-JUL-1999;	99US-0145085.
PR		PR	22-JUL-1999;	99US-0145087.
PR		PR	22-JUL-1999;	99US-0145089.
PR		PR	22-JUL-1999;	99US-0145192.
PR		PR	23-JUL-1999;	99US-0145145.
PR		PR	23-JUL-1999;	99US-0145218.
PR		PR	23-JUL-1999;	99US-0145224.
PR		PR	26-JUL-1999;	99US-0145276.
PR		PR	27-JUL-1999;	99US-0145913.
PR		PR	27-JUL-1999;	99US-0145918.
PR		PR	27-JUL-1999;	99US-0145919.
PR		PR	28-JUL-1999;	99US-0145951.
PR		PR	02-AUG-1999;	99US-0146386.
PR		PR	02-AUG-1999;	99US-0146388.
PR		PR	02-AUG-1999;	99US-0146389.
PR		PR	03-AUG-1999;	99US-0147038.
PR		PR	04-AUG-1999;	99US-0147204.
PR		PR	04-AUG-1999;	99US-0147302.
PR		PR	05-AUG-1999;	99US-0147192.
PR		PR	05-AUG-1999;	99US-0147260.
PR		PR	06-AUG-1999;	99US-0147303.
PR		PR	06-AUG-1999;	99US-0147416.
PR		PR	09-AUG-1999;	99US-0147493.
PR		PR	09-AUG-1999;	99US-0147935.
PR		PR	10-AUG-1999;	99US-0148171.
PR		PR	11-AUG-1999;	99US-0148319.
PR		PR	12-AUG-1999;	99US-0148341.
PR		PR	13-AUG-1999;	99US-0148565.
PR		PR	13-AUG-1999;	99US-0148684.
PR		PR	16-AUG-1999;	99US-0149368.
PR		PR	17-AUG-1999;	99US-0149175.
PR		PR	18-AUG-1999;	99US-0149426.
PR		PR	20-AUG-1999;	99US-0149722.
PR		PR	20-AUG-1999;	99US-0149723.
PR		PR	20-AUG-1999;	99US-0149929.
PR		PR	23-AUG-1999;	99US-0149902.
PR		PR	23-AUG-1999;	99US-0149930.
PR		PR	25-AUG-1999;	99US-0150566.
PR		PR	26-AUG-1999;	99US-0150884.
PR		PR	27-AUG-1999;	99US-0151065.
PR		PR	27-AUG-1999;	99US-0151066.
PR		PR	27-AUG-1999;	99US-0151080.
PR		PR	30-AUG-1999;	99US-0151303.
PR		PR	31-AUG-1999;	99US-0151438.
PR		PR	01-SEP-1999;	99US-0151930.
PR		PR	07-SEP-1999;	99US-0152363.
PR		PR	10-SEP-1999;	99US-0153070.
PR		PR	13-SEP-1999;	99US-0153758.
PR		PR	15-SEP-1999;	99US-0154018.
PR		PR	16-SEP-1999;	99US-0154039.
PR		PR	20-SEP-1999;	99US-0154779.



PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147304.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 03-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 43.3%; Score 45; DB 21; Length 242;

Best Local Similarity 44.4%; Pred. No. 7.4;

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 NFPHGMLDLEETAANSKD 20

Db 68 dypvgllldiekilppgkd 85

RESULT 12

GI0096

ID GI0096 standard; Protein; 383 AA.

XX AC GI0096;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 8282.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0129845.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.

```

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.
PR 29-OCT-1999; 99US-0162144.

Query Match 43.3%; Score 45; DB 21; Length 383;
Best Local Similarity 44.4%; Pred. No. 12;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 NFPHGMLDEEIAANSKD 20
   :|:|:|:|:|
Db 209 dypvgildiekilppgkd 226

RESULT 13
G32339
ID G32339 standard; Protein: 383 AA.
XX
AC G32339;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38994.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

```

```
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.3%; Score 45; DB 21; Length 383;
Best Local Similarity 44.4%; Pred. No. 12;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 NPHGMLDLEEIAANSKD 20
   :| |::|:| |
Db 209 dypvgllidiekilppgkd 226

RESULT 14
G10095
ID G10095 standard; Protein; 399 AA.
XX
AC G10095;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8281.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
```



PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0140354.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:35:04 ; Search time 62.1 Seconds  
(without alignments)  
6.187 Million cell updates/sec

Title: US-09-171-432a-44

Perfect score: 104  
Sequence: 1 KVNFPFGMLDLEIAANSKD 20

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210957 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	2227	4 US-08-475-886-2	Sequence 2, Appli
2	104	100.0	2227	4 US-08-475-886-4	Sequence 4, Appli
3	104	100.0	2227	4 US-08-475-886-6	Sequence 6, Appli
4	104	100.0	2227	4 US-08-397-232-2	Sequence 2, Appli
5	104	100.0	2227	4 US-08-397-232-4	Sequence 4, Appli
6	45	43.3	417	2 US-09-099-677A-6	Sequence 6, Appli
7	45	43.3	417	3 US-09-261-471-6	Sequence 6, Appli
8	44	42.3	432	2 US-09-099-677A-3	Sequence 3, Appli
9	44	42.3	432	3 US-09-261-471-3	Sequence 3, Appli
10	43	41.3	98	4 US-09-181-487-2	Sequence 2, Appli
11	39	37.5	579	1 US-08-126-564A-31	Sequence 31, Appli
12	39	37.5	579	5 PCT-US94-09143-31	Sequence 31, Appli
13	38	36.5	1250	4 US-08-938-291A-9	Sequence 9, Appli
14	38	36.5	1367	2 US-08-249-687C-2	Sequence 2, Appli
15	38	36.5	1367	2 US-08-625-819-2	Sequence 2, Appli
16	38	36.5	1367	3 US-08-746-559A-2	Sequence 2, Appli
17	37.5	36.1	568	4 US-09-238-373-2	Sequence 2, Appli
18	37	35.6	823	1 US-08-261-465-1	Sequence 1, Appli
19	37	35.6	823	1 US-08-405-254-5	Sequence 5, Appli
20	36.5	35.1	350	2 US-08-828-922-1	Sequence 1, Appli
21	36.5	35.1	4344	1 US-08-469-486-52	Sequence 52, Appli
22	36.5	35.1	4544	1 US-08-469-658-52	Sequence 52, Appli
23	36	34.6	95	3 US-08-946-329A-86	Sequence 86, Appli
24	36	34.6	434	2 US-08-823-516-145	Sequence 145, Appli
25	36	34.6	498	2 US-08-511-485-13	Sequence 13, Appli
26	36	34.6	793	2 US-08-468-558-5	Sequence 5, Appli
27	36	34.6	2004	1 US-08-375-709-15	Sequence 15, Appli

Sequence 15, Appli  
Sequence 9, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 10, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 6, Appli  
Sequence 5, Appli  
Sequence 8, Appli  
Sequence 2, Appli

28 36 34.6 2004 1 US-08-752-929-15  
29 36 34.6 2004 4 US-09-090-793-9  
30 35 33.7 108 2 US-08-825-782-1  
31 35 33.7 246 1 US-08-286-888B-1  
32 35 33.7 246 1 US-08-294-386C-1  
33 35 33.7 246 1 US-08-294-386C-3  
34 35 33.7 246 1 US-08-299-249A-1  
35 35 33.7 246 1 US-08-299-249A-10  
36 35 33.7 246 1 US-08-590-708-1  
37 35 33.7 246 5 PCT-US94-06543-1  
38 35 33.7 246 5 PCT-US95-10224-1  
39 35 33.7 246 5 PCT-US95-10224-3  
40 35 33.7 292 1 US-08-286-888B-6  
41 35 33.7 347 1 US-07-637-870-5  
42 35 33.7 347 1 US-07-640-476-10  
43 35 33.7 348 1 US-07-637-399-8  
44 35 33.7 348 1 US-08-112-703-8  
45 35 33.7 374 1 US-08-625-876-2

## ALIGNMENTS

RESULT 1  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2

Query Match 100.0%; Score 104; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEIAANSKD 20  
Db 922 KVNFPFGMLDLEIAANSKD 941

RESULT 2  
US-08-475-886-4  
; Sequence 4, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18

```
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match      100.0%; Score 104; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVNPHGMLDLEETAANSKD 20
Db 922 KVNPHGMLDLEETAANSKD 941

RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match      100.0%; Score 104; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVNPHGMLDLEETAANSKD 20
Db 922 KVNPHGMLDLEETAANSKD 941

RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: PatentIn Ver. 2.1
US-08-397-232-2

Query Match      100.0%; Score 104; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVNPHGMLDLEETAANSKD 20
Db 922 KVNPHGMLDLEETAANSKD 941

RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

Query Match      100.0%; Score 104; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVNPHGMLDLEETAANSKD 20
Db 922 KVNPHGMLDLEETAANSKD 941

RESULT 6
US-09-099-677A-6
; Sequence 6, Application US/09099677A
; Patent No. 5965369
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN SUCCINYL-COENZYME A SYNTHETASE HOLOENZYME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/099,677A
; FILING DATE: June 18, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0545 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: g164669
; US-09-099-677A-6

Query Match 43.3%; Score 45; DB 2; Length 417;
Best Local Similarity 42.1%; Pred. No. 6.3;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEEIAANSKD 20
Db 153 VGSPQGGVDIEEVAASNP 171

RESULT 7
US-09-261-471-6
; Sequence 6, Application US/09261471
; Patent No. 6025123
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN SUCCINYL-COA SYNTHETASE HOLOENZYME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; FILING DATE: June 18, 1998
; APPLICATION NUMBER: US/09/261,471
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0545 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: g164669
; US-09-099-677A-6

Query Match 43.3%; Score 45; DB 2; Length 417;
Best Local Similarity 42.1%; Pred. No. 6.3;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEEIAANSKD 20
Db 153 VGSPQGGVDIEEVAASNP 171

RESULT 7
US-09-261-471-6
; Sequence 6, Application US/09261471
; Patent No. 6025123
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN SUCCINYL-COA SYNTHETASE HOLOENZYME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; FILING DATE: June 18, 1998
; APPLICATION NUMBER: US/09/261,471
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0545 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: g164669
; US-09-099-677A-6
```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: g164669
; US-09-261-471-6

Query Match 43.3%; Score 45; DB 3; Length 417;
Best Local Similarity 42.1%; Pred. No. 6.3;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEEIAANSKD 20
Db 153 VGSPQGGVDIEEVAASNP 171

RESULT 8
US-09-099-677A-3
; Sequence 3, Application US/09099677A
; Patent No. 5965369
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN SUCCINYL-COENZYME A SYNTHETASE HOLOENZYME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/099,677A
; FILING DATE: June 18, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0545 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSBPT06
; CLONE: 3273853
; US-09-099-677A-3

Query Match 42.3%; Score 44; DB 2; Length 432;
Best Local Similarity 42.1%; Pred. No. 9.7;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEEIAANSKD 20
Db 168 VGPOQGGVDIEEVAASNP 186
```

```

RESULT 11
US-8-126-564A-31
: Sequence 31, Application US/08126564A
: Patent No. 5436150
: GENERAL INFORMATION:
: APPLICANT: Chandrasegaran, Srinivasan
: TITLE OF INVENTION: Functional Domains in FokI
: TITLE OF INVENTION: Restriction Endonuclease
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cushman, Darby & Cushman
: STREET: 1100 New York Ave., N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3918
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0,
: SOFTWARE: Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/126,564A
: FILING DATE: 27-SEPTEMBER-93
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Kokulis, Paul N.
: REGISTRATION NUMBER: 16,773
: REFERENCE/DOCKET NUMBER: PNK/41130/82506/CLB
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-861-3503
: TELEFAX: 202-822-0944

```



TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 579 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-126-564A-31

Query Match 37.5%; Score 39; DB 1; Length 579;  
Best Local Similarity 57.1%; Pred. No. 95;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 PHGMLDLEETAANS 18  
|| :||| ||  
Db 405 PHEYLIEIETARN 418

RESULT 12  
PCT-US94-09143-31  
Sequence 31, Application PC/TUS9409143  
GENERAL INFORMATION:  
APPLICANT: Chandrasegaran, Srinivasan  
TITLE OF INVENTION: Functional Domains in Foki  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cushman, Darby & Cushman  
STREET: 1100 New York Ave., N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0,  
SOFTWARE: Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/09143  
FILING DATE: 23-AUG-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/126,564  
FILING DATE: 27-SEPTEMBER-93  
ATTORNEY/AGENT INFORMATION:  
NAME: Kokulis, Paul N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: PNK/4130/82506/CLB  
TELEPHONE: 202-861-3503  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 579 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US94-09143-31

Query Match 37.5%; Score 39; DB 5; Length 579;  
Best Local Similarity 57.1%; Pred. No. 95;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 PHGMLDLEETAANS 18  
|| :||| ||  
Db 405 PHEYLIEIETARN 418

RESULT 13  
US-08-938-291A-9  
Sequence 9, Application US/08938291A  
Patent No. 6117673  
GENERAL INFORMATION:  
APPLICANT: Lev, Sima  
APPLICANT: Plowman, Gregory D.  
APPLICANT: Schlensing, Joseph  
TITLE OF INVENTION: RDCB PROTEINS AND RELATED  
PRODUCTS AND METHODS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,291A  
FILING DATE: September 26, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,337  
FILING DATE: October 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 228/172  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1250 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-938-291A-9

Query Match 36.5%; Score 38; DB 4; Length 1250;  
Best Local Similarity 45.5%; Pred. No. 3 4e+02;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NFPHGMLDLEE 13  
|||||:  
Db 1120 NFPHGLISFAD 1130

RESULT 14  
US-08-249-687C-2  
Sequence 2, Application US/08249687C  
Patent No. 5942412  
GENERAL INFORMATION:  
APPLICANT: PRAGER, DIANE  
APPLICANT: MELMED, SHLOMO  
TITLE OF INVENTION: POLYNUCLEIC ACID ENCODING  
VARIANT INSULIN-LIKE GROWTH FACTOR I RECEPTOR BETA  
SUBUNIT & RECEPTOR  
TITLE OF INVENTION: SUBUNIT & RECEPTOR  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder & Poplawski  
STREET: 444 South Flower St., 19th Floor  
CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/249,687C  
FILING DATE: 26-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/044,540  
FILING DATE: 06-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Basile, Lena  
REGISTRATION NUMBER: P-44,026  
REFERENCE/DOCKET NUMBER: P07 32349  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213-622-7700  
TELEFAX: 213-489-4210  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1367 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-249-687C-2

Query Match 36.5%; Score 38; DB 2; Length 1367;  
Best Local Similarity 37.5%; Pred. No. 3.8e+02;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 FPHGMLDLEETAAANSK 19  
: | :||: | |  
Db 673 YADGTIDIEEVTEPNK 688

RESULT 15  
US-08-625-819-2  
; Sequence 2, Application US/08625819  
; Patent No. 5958872  
; GENERAL INFORMATION:  
; APPLICANT: O'CONNOR, Rosemary; and  
; APPLICANT: BASERGA, Renato L.  
; TITLE OF INVENTION: ACTIVE SURVIVAL DOMAINS OF IGF-IR  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE and DORR LLP  
; STREET: 1455 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/625,819  
; FILING DATE: 01-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WIXON, Henry N.

REGISTRATION NUMBER: 32,073  
REFERENCE/DOCKET NUMBER: 104322.162  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 942-8459  
TELEFAX: (202) 942-8484  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1367 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-625-819-2

Query Match 36.5%; Score 38; DB 2; Length 1367;  
Best Local Similarity 37.5%; Pred. No. 3.8e+02;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 FPHGMLDLEETAAANSK 19  
: | :||: | |  
Db 673 YADGTIDIEEVTEPNK 688

Search completed: April 24, 2001, 16:35:05  
Job time: 536 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:25:57 ; Search time 73.14 seconds.  
(without alignments)  
18.792 Million cell updates/sec

Title: US-09-171-432A-44

Perfect score: 104

Sequence: 1 KVNFPHGMLDLEEAANSKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_67:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	2227	1 GNNYHM	genome polyprotein
2	104	100.0	2227	1 GNNYHR	genome polyprotein
3	104	100.0	2227	1 GNNYMK	genome polyprotein
4	104	100.0	2227	1 GNNYHB	genome polyprotein
5	101	97.1	2230	1 GNNYSA	genome polyprotein
6	47	45.2	506	1 S58522	glycine--trna liga
7	47	45.2	1079	2 T30996	hypothetical prote
8	46	44.2	259	2 F82087	deoxyribose-phosph
9	46	44.2	677	2 H64574	DNA topoisomerase
10	45	43.3	195	2 T08812	probable succinate
11	45	43.3	417	2 A44529	succinate--CoA lig
12	45	43.3	852	2 B72685	hypothetical prote
13	45	43.3	935	2 T19011	hypothetical prote
14	45	43.3	1213	2 T19835	hypothetical prote
15	44	42.3	223	2 T37962	hypothetical prote
16	44	42.3	500	2 S50508	ANP1 protein - yea
17	43.5	41.8	373	2 A63773	hypothetical prote
18	43	41.3	140	2 E81659	conserved hypothet
19	43	41.3	141	2 G71501	hypothetical prote
20	43	41.3	309	2 H71089	hypothetical prote
21	43	41.3	327	2 T00876	hypothetical prote
22	43	41.3	440	2 B71858	adenylsuccinate l
23	43	41.3	440	2 H64658	adenylsuccinate l
24	43	41.3	917	2 T21870	hypothetical prote
25	42	40.4	222	2 C82343	conserved hypothet
26	42	40.4	229	2 S77449	hydrogenase expres
27	42	40.4	290	2 S19426	hypothetical prote
28	42	40.4	309	2 T33259	hypothetical prote
29	42	40.4	342	2 C72313	hypothetical prote

30	42	40.4	365	2 B48945	recombination prot
31	42	40.4	384	2 E82088	conserved hypothet
32	42	40.4	495	2 D64578	conserved hypothet
33	42	40.4	539	2 G70520	probable csp prote
34	42	40.4	1742	2 S24600	projectin - fruit
35	42	40.4	6658	2 T13931	projectin - fruit
36	41	39.4	193	2 G64187	anthranilate synth
37	41	39.4	218	2 D71693	hypothetical prote
38	41	39.4	274	2 T36347	hypothetical prote
39	41	39.4	365	2 T20652	hypothetical prote
40	41	39.4	400	2 C34443	nitrogenase cofact
41	41	39.4	607	2 B84153	two-component sens
42	41	39.4	834	2 T19010	hypothetical prote
43	40.5	38.9	237	1 B8AG58	virB8 protein - Ag
44	40.5	38.9	434	2 E71638	UDP-glucose 6-dehy
45	40.5	38.9	1286	1 H36845	DNA-directed RNA p

#### ALIGNMENTS

RESULT 1

GNNYHM

genome polyprotein - human hepatitis A virus (strain HM-175, wild type)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
B: RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D

C:Species: human hepatitis A virus

A:Note: host Homo sapiens (man)

C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C:Accession: A25981

R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.

J. Virol. 61, 50-59, 1987

A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with  
A:Reference number: A25981; MUID:87061253

A:Accession: A25981

A:Molecule type: genomic RNA

A:Residues: 1-2227 <COH>

A:Cross-References: EMBL:M14707; NID:g329582; PIDN:AAA5465.1; PID:g329583

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>

F:24-245/Product: coat protein 1B #status predicted <VP2>

F:246-491/Product: coat protein 1C #status predicted <VP3>

F:492-980/Product: coat protein 1D #status predicted <VP1>

F:792-980/Product: core protein 2A #status predicted <C2A>

F:981-1087/Product: core protein 2B #status predicted <C2B>

F:1088-1422/Product: core protein 2C #status predicted <C2C>

F:1423-1496/Product: protein 3A #status predicted <C3A>

F:1497-1519/Product: protein 3B #status predicted <C3B>

F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>

F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 104; DB 1; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KVNFPHGMLDLEEAANSKD 20

|||||

Db 922 KVNFPHGMLDLEEAANSKD 941

RESULT 2

GNNYHR

genome polyprotein - human hepatitis A virus

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core  
NA polymerase (EC 2.7.7.48), protein 3D

C:Species: human hepatitis A virus

A:Note: host Homo sapiens (man)

C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999

C:Accession: A03903

R:Najarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van N  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985



A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 97.1%; Score 101; DB 1; Length 2230;  
Best Local Similarity 95.0%; Pred. No. 3.2e-08;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEIAANSKD 20  
|||||:|||||:|||||  
Db 926 KVNFPFGMLDLEIAANSKD 945

RESULT 6  
S58522  
glycine--trna ligase (EC 6.1.1.14) - Thermus aquaticus  
N:Alternate names: glycyl-trna synthetase  
C:Species: Thermus aquaticus  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S58522  
R:Logan, D.T.; Mazauric, M.H.; Kern, D.; Moras, D.  
EMBO J. 14, 4156-4167, 1995  
A:Title: Crystal structure of glycyl-trna synthetase from Thermus thermophilus.  
A:Reference number: S58522; MUID:96016187  
A:Accession: S58522  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-506 <LOG>  
A:Experimental source: strain HB8  
A:Note: the source is designated as Thermus thermophilus  
C:Superfamily: Mycoplasma genitalium glycine--trna ligase  
C:Keywords: aminoacyl-trna synthetase; ATP; ligase; protein biosynthesis

Query Match 45.2%; Score 47; DB 1; Length 506;  
Best Local Similarity 75.0%; Pred. No. 6.4;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 FPHGMLDLEIA 15  
|||||:|||||  
Db 299 FPHGSELEGIA 310

RESULT 7  
T30996  
hypothetical protein C01G5.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T30996  
R:Bradshaw, H.; Stellyes, L.  
submitted to the EMBL Data Library, August 1999  
A:Description: The sequence of C. elegans cosmid C01G5.  
A:Reference number: 220956  
A:Accession: T30996  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1079 <BRA>  
A:Cross-references: EMBL:U50068; PIDN:AAB37736.2  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Map position: IV

A:Introns: 10/1; 31/1; 75/1; 108/1; 144/1; 927/2; 1056/2  
A:Note: C01G5.4

Query Match 45.2%; Score 47; DB 2; Length 1079;  
Best Local Similarity 36.8%; Pred. No. 15;  
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEIAANSKD 20  
|:|||||:|:|  
Db 410 IGFSRGILDLDKHVAGDARD 428

RESULT 8  
F82087  
deoxyribose-phosphate aldolase VC2350 [imported] - Vibrio cholerae (group O1 strain N  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 17-Nov-2000  
C:Accession: F82087  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: F82087  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <HEI>  
A:Cross-references: GB:AE003852; GB:AE003852; NID:g9656912; PIDN:AAF95493.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2350  
A:Map position: 1  
C:Superfamily: deoxyribose-phosphate aldolase

Query Match 44.2%; Score 46; DB 2; Length 259;  
Best Local Similarity 52.9%; Pred. No. 4.3;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFPFGMLDLEIAANSK 19  
|||||:|:|:|  
Db 75 NFPFGNDDEIAVAETK 91

RESULT 9  
H64574  
DNA topoisomerase I - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: H64574  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467  
A:Accession: H64574  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-677 <TOM>  
A:Cross-references: GB:AE000559; GB:AE000511; NID:g2313536; PIDN:AAD07502.1; PID:g231  
C:Superfamily: DNA topoisomerase I

Query Match 44.2%; Score 46; DB 2; Length 677;  
Best Local Similarity 31.6%; Pred. No. 13;  
Matches 6; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEIAANSK 19  
::||:||||:|:|

Db 346 RITHPHALKDLEKVCSDAK 364

RESULT 10

T08812

probable succinate-CoA ligase (GDP-forming) (EC 6.2.1.4) beta chain - human (fragment)

N:Alternate names: protein DKFZp586M2023.1; succinyl-CoA synthetase (GDP-forming) beta c

C:Species: Homo sapiens (man)

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 05-May-2000

C:Accession: T08812

R:Ansorge, W.; Wirkner, U.; Meves, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16472

A:Accession: T08812

A:Molecule type: mRNA

A:Residues: 1-195 <ANS>

A:Cross-references: EMBL:AL050226

A:Experimental source: adult uterus; clone DKFZp586M2023

C:Genetics:

A:Note: DKFZp586M2023.1

C:Function:

A:Description: catalyzes the formation of succinyl-CoA from succinate with concomitant h

C:Superfamily: succinate-CoA ligase (ADP-forming) beta chain

C:Keywords: acid-thiol ligase; coenzyme A; mitochondrion

Query Match 43.3%; Score 45; DB 2; Length 195;

Best Local Similarity 42.1%; Pred. No. 4.6;

Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEIEIAANSKD 20

|||:|:|:|:|:|

Db 65 VGSGPGGVDIIEVAASNP 83

RESULT 11

A44529

succinate-CoA ligase (GDP-forming) (EC 6.2.1.4) beta chain precursor - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 30-Sep-1993 #sequence\_revision 07-Apr-1994 #text\_change 05-May-2000

C:Accession: A44529

R:Baily, D.L.; Wolodko, W.T.; Bridger, W.A.

Protein Sci. 2, 1255-1262, 1993

A:Title: Cloning, characterization and expression of the beta subunit of pig heart succi

A:Reference number: A44529; MUID:94004462

A:Accession: A44529

A:Molecule type: mRNA

A:Residues: 1-417 <BAI>

A:Cross-references: GB:L06944

C:Genetics:

A:Start codon: CTG

C:Function:

A:Description: catalyzes the formation of succinyl-CoA from succinate with concomitant h

C:Superfamily: succinate-CoA ligase (ADP-forming) beta chain

C:Keywords: acid-thiol ligase; coenzyme A; heterodimer; mitochondrion

F:1-23/Domain: transit peptide (mitochondrion) (fragment) #status predicted <TNP>

F:24-417/Product: succinate-CoA ligase (GDP-forming) (EC 6.2.1.4) beta chain #status pr

Query Match 43.3%; Score 45; DB 2; Length 417;

Best Local Similarity 42.1%; Pred. No. 11;

Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEIEIAANSKD 20

|||:|:|:|:|:|

Db 153 VGSGPGGVDIIEVAASNP 171

RESULT 12

B72685

hypothetical protein APE0898 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C:Accession: B72685

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339

A:Accession: B72685

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-852 <KAW>

A:Cross-references: DDBJ:AF000060; NID:g5104188; PIDN:BAA79882.1; PID:g5104567

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0898

C:Superfamily: Aeropyrum pernix hypothetical protein APE0898

Query Match 43.3%; Score 45; DB 2; Length 852;

Best Local Similarity 41.2%; Pred. No. 25;

Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEIEIAANS 18

|||:|:|:|:|:|

Db 355 VNGPYGLIDISDIYVDS 371

RESULT 13

T19011

hypothetical protein C06C6.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T19011

R:McMurray, A.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19059

A:Accession: T19011

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-935 <WIL>

A:Cross-references: EMBL:Z93374; PIDN:CAB07557.1; GSPDB:GN00023; CESP:C06C6.7

A:Experimental source: clone C06C6

C:Genetics:

A:Gene: CESP:C06C6.7

A:Map position: 5

A:Introns: 28/1; 55/1; 80/1; 801/2; 865/2

Query Match 43.3%; Score 45; DB 2; Length 935;

Best Local Similarity 41.2%; Pred. No. 28;

Matches 7; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

QY 4 FPHGMLDLEIEIAANSKD 20

|||:|:|:|:|:|

Db 324 FANGISDIEKLSSDSKD 340

RESULT 14

T19835

hypothetical protein C38D9.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T19835; T27344

R:Ainscough, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19184

A:Accession: T19835

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1213 <WIL>

A:Cross-references: EMBL:Z81481; PIDN:CAB03951.1; GSPDB:GN00023; CESP:C38D9.5

A:Experimental source: clone C38D9

R:Lloyd, C.

submitted to the EMBL Data Library, March 1997



A:Reference number: Z20348  
A:Accession: T27344  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1213 <W12>  
A:Cross-references: EMBL:Z92973; PIDN:CAB07494.1; GSPDB:GN00023; CESP:C38D9.5  
C:Genetics:  
A:Map position: 5  
A:Gene: CESP:C38D9.5  
A:Introns: 32/1; 490/3; 672/1; 770/2; 804/1; 863/1; 967/3; 1000/2; 1072/1

Query Match 43.3%; Score 45; DB 2; Length 1213;  
Best Local Similarity 44.4%; Pred. No. 38;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEETANSK 19  
DB 312 LGFPHGVSDILQIQENLK 329

## RESULT 15

T37962  
hypothetical protein SPAC19D5.02c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C:Accession: T37962  
R:Devlin, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z21757  
A:Accession: T37962  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-223 <DEV>  
A:Cross-references: EMBL:Z99531; PIDN:CAB16712.1; GSPDB:GN00066; SPDB:SPAC19D5.02c  
C:Genetics:  
A:Map position: 1  
A:Gene: SPDB:SPAC19D5.02c  
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC19D5.02c

Query Match 42.3%; Score 44; DB 2; Length 223;  
Best Local Similarity 41.2%; Pred. No. 7.9;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 4 FPHGMLDLEETANSKD 20  
DB 35 FPHGTIDIPETFRPSD 51

Search completed: April 24, 2001, 16:25:59  
Job time: 293 sec

**THIS PAGE BLANK (uspto)**

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:34:03 ; Search time 38.74 Seconds  
(without alignments)  
17.685 Million cell updates/sec

Title: US-09-171-432A-44

Perfect score: 104

Sequence: 1 KVNFPFGMLDLEETAANSKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	2226	1 POLG_HPAVA	P26581 hepatitis a
2	104	100.0	2226	1 POLG_HPAV8	P26582 hepatitis a
3	104	100.0	2227	1 POLG_HPAVH	P08617 hepatitis a
4	104	100.0	2227	1 POLG_HPAVL	P06441 hepatitis a
5	104	100.0	2227	1 POLG_HPAVM	P13901 hepatitis a
6	101	97.1	2230	1 POLG_HPAVS	P14553 simian hepa
7	100	96.2	2226	1 POLG_HPAV2	P26580 hepatitis a
8	47	45.2	505	1 SYG_THETH	P58206 thermus aqu
9	45	43.3	417	1 SUCB_PIG	P53590 sus scrofa
10	44	42.3	500	1 ANPL_YEAST	P32629 saccharomyc
11	43	41.3	98	1 SRG1_HUMAN	O75711 homo sapien
12	43	41.3	357	1 UB2_CHICK	O74229 gallus gall
13	43	41.3	440	1 PUR8_HELPJ	O92ka2 helicobacte
14	43	41.3	440	1 PUR8_HELPY	P58468 helicobacte
15	42	40.4	214	1 GLP1_PHANI	P25853 pharbitis n
16	42	40.4	290	1 YC06_YEAST	P25617 saccharomyc
17	42	40.4	365	1 REC2_LACLA	Q01840 lactococcus
18	41	39.4	193	1 YB71_HAETIN	P44339 haemophilus
19	41	39.4	218	1 Y363_RICPR	O92d96 rickettsia
20	41	39.4	400	1 NIFS_ANAAZ	Q43884 anabaena az
21	41	39.4	400	1 NIFS_ANASP	P25623 anabaena sp
22	40.5	38.9	237	1 VIB8_AGR75	P17798 agrobacteri
23	40.5	38.9	434	1 UDG_RICPR	O05973 rickettsia
24	40.5	38.9	1286	1 RP01_VACCC	P20504 vaccinia vi
25	40.5	38.9	1286	1 RP01_VARV	P33053 variola vir
26	40.5	38.9	1287	1 RP01_VACCV	P07392 vaccinia vi
27	40	38.5	310	1 YCCK_BACSU	P46905 bacillus su
28	40	38.5	353	1 UB2_HUMAN	O75604 homo sapien
29	40	38.5	353	1 UB2_MOUSE	O88623 mus musculu
30	40	38.5	446	1 SYG_MWCGE	P47493 mycoplasma
31	40	38.5	507	1 YIK4_YEAST	P40486 saccharomyc
32	40	38.5	558	1 PILF_NEIGO	P37094 neisseria g
33	40	38.5	646	1 KDBE_SCHPO	Q10364 schizosacch

## RESULT 1

ID	POLG_HPAVA	STANDARD;	PRT;	2226 AA.
AC	P26581;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)]			
OS	Hepatitis A virus (strain 43c).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;			
OC	Hepatovirus.			
OX	NCBI_TaxID=12095;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91162758; PubMed=1705995;			
RA	Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,			
RA	Cromeans T., Jansen R.W.;			
RT	"Antigenic and genetic variation in cytopathic hepatitis A virus			
RT	variants arising during persistent infection: evidence for genetic			
RT	recombination."			
RL	J. Virol. 65:2056-2065(1991).			
CC	!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS.			
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,			
CC	VP3, AND VP4.			
CC	!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.			
CC	!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M59809; AAA45469.1; -			
DR	MEROPS; C03.005; -			
DR	InterPro; IPR001205; -			
DR	InterPro; IPR001205; -			
DR	Pfam; PF00680; RNA_dep_RNA_pol; 1.			
DR	Pfam; PF00910; RNA_helicase; 1.			
KW	Polyprotein; Coat protein; Core protein; Transferase;			
KW	RNA-directed RNA polymerase; Hydrolase; Thiol protease.			
FT	CHAIN 1 23			
FT	CHAIN 24 245			
FT	CHAIN 246 491			
FT	CHAIN 492 794			
FT	CHAIN 795 900			
FT	CHAIN 901 1087			
FT	CHAIN 1088 1422			
FT	CHAIN 1423 1495			
FT	CHAIN 1496 1518			
FT	CHAIN 1519 1737			
FT	CHAIN 1738 2226			

P00882 escherichia  
Q38175 methanococc  
P75174 mycoplasma  
P14870 flavobacter  
Q05085 arabidopsis  
P17169 e glucosami  
P57332 buchnera ap  
Q48253 helicobacte  
P45444 emericella  
P55528 rhizobium s  
Q10824 mycobacteri  
P06172 xenopus lae

## ALIGNMENTS

```
Query Match 100.0%; Score 104; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 1.le-09;
```

```

DR MEROPS; C03.005; -
DR InterPro; IPR000605; -
DR InterPro; IPR001205; -
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 836
FT CHAIN 837 980
FT CHAIN 981 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1496
FT CHAIN 1497 1519
FT CHAIN 1520 1738
FT CHAIN 1739 2227
FT CHAIN 2227 77
FT VARIANT 784 764
FT VARIANT 821 821
FT VARIANT 1052 1052
FT VARIANT 1062 1062
FT VARIANT 1118 1118
FT VARIANT 1151 1151
FT VARIANT 1163 1163
FT VARIANT 1277 1277
FT VARIANT 1500 1500
FT VARIANT 1805 1805
FT VARIANT 1930 1930
FT VARIANT 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;

Query Match 100.0%; Score 104; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVNFPFGMLDLEEIAANSKD 20
DB 922 KVNFPFGMLDLEEIAANSKD 941
|||||
RESULT 5
POLG_HPAPV STANDARD; PRT; 2227 AA.
AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;
AC Q81090; Q81091; Q81092; Q81093;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain MBB).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85190549; PubMed=2986127;
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,
RA Merryweather J., van Nest G., Dina D.;
RT "Primary structure and gene organization of human hepatitis A virus.";
RT Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



4 FPHGMLDLEEIA 15  
||||| I:||||  
298 FPHGSLELEGIA 309

LT 9

PIG  
SUBC\_PIG STANDARD; PRT; 417 AA.  
P53590; Q95279;  
01-OCT-1996 (Rel. 34, Created)  
01-OCT-1996 (Rel. 34, Last sequence update)  
01-OCT-2000 (Rel. 40, Last annotation update)  
SUCCINYL-COA LIGASE [GDP-FORMING] BETA-CHAIN, MITOCHONDRIAL PRECURSOR  
(EC 6.2.1.14) (SUCCINYL-COA SYNTHETASE, BETA CHAIN) (SCS-BETA).  
SUCUGL2.  
Sus scrofa (Pig).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
NCBI\_TaxID=9823;  
[1]  
SEQUENCE FROM N.A., AND SEQUENCE OF 23-41.  
TISSUE=Heart;  
MEDLINE=94004462; PubMed=8401211;  
Bailey D.L., Wolodko W.T., Bridger W.A.;  
"Cloning, characterization, and expression of the beta subunit of pig  
heart succinyl-CoA synthetase.";  
Protein Sci. 2:1255-1262(1993).  
[2]  
SEQUENCE OF 1-82 FROM N.A.  
TISSUE=Small intestine;  
MEDLINE=96327607; PubMed=8672129;  
Winteroe A.K., Fredholm M., Davies W.;  
"Evaluation and characterization of a porcine small intestine cDNA  
library: analysis of 839 clones.";  
Mamm. Genome 7:509-517(1996).  
-!- CATALYTIC ACTIVITY: SUCCINATE + COA + GTP -> SUCCINYL-COA + GDP +  
ORTHOPHOSPHATE.  
-!- PATHWAY: SUBSTRATE LEVEL PHOSPHORYLATION STEP OF THE TRICARBOXYLIC  
ACID CYCLE.





RL J. Biol. Chem. 273:18015-18018(1998).

CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).

CC -1- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY IN THE CENTRAL NERVOUS

CC SYSTEM OF ADULT, BUT NOT AT ALL IN FETAL BRAIN. HIGH LEVELS OF

CC SCRG1 TRANSCRIPTS ARE ALSO OBSERVED IN TESTIS AND AORTA.

CC

CC -1- SIMILARITY: BELONGS TO THE SCRG1 FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL: AJ224677; CAAL2059.1; .

DR HSSP: P01544; 2PLH.

DR MIM: 603163; .

KW SIGNAL.

DR SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 98 SCRAPIE-RESPONSIVE PROTEIN 1.

FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 98 AA; 11081 MW; A5F1FD40BF5401C0 CRC64;

Query Match 41.3%; Score 43; DB 1; Length 98;

Best Local Similarity 44.4%; Pred. No. 1.4;

Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 3 NFPHGMLDLEEIAANSKD 20

DB 40 NLPEGVADLTQIDNVNVD 57

RESULT 12

UBP2\_CHICK

ID UB2\_CHICK STANDARD; PRT; 357 AA.

AC Q37429;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 2 (EC 3.1.2.15) (UBIQUITIN

DE THIOLESTERASE 2) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 2)

DE (UBIQUITIN-DEGRADATING ENZYME 2) (41 KDA UBIQUITIN-SPECIFIC PROTEASE).

OS USP2 OR UBPA1.

GN Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RT TISSUE=Skeletal muscle;

RX MEDLINE=97467343; PubMed=9325273;

RA Baek S., Choi K.S., Yoo Y.J., Cho J.M., Baker R.T., Tanaka K.,

RA Chung C.H.;

RA "Molecular cloning of a novel ubiquitin-specific protease, UBPA1, with

RT isopeptidase activity in chick skeletal muscle.";

RL J. Biol. Chem. 272:25560-25565(1997).

CC -1- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =

CC UBIQUITIN + A THIOL.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS

CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL: AF016107; AAC13729.1; .

DR MEROPS; C19.013; .

DR InterPro: IPR001394; .

DR Pfam: PF00442; UCH\_1; 1.

DR Pfam: PF00443; UCH\_2; 1.

DR PROSITE: PS00972; UCH\_2.1; 1.

DR PROSITE: PS00973; UCH\_2.2; 1.

DR PROSITE: PS00975; UCH\_2.3; 1.

DR Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family.

FT ACT\_SITE 28 28 BY SIMILARITY.

FT ACT\_SITE 301 301 BY SIMILARITY.

FT ACT\_SITE 309 309 BY SIMILARITY.

SQ SEQUENCE 357 AA; 40931 MW; EC39E6454937C455 CRC64;

Query Match 41.3%; Score 43; DB 1; Length 357;

Best Local Similarity 58.8%; Pred. No. 6.2;

Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 VNFPHGMLDLEEIAANS 18

DB 271 VNFPLKDLDFEFASQS 287

RESULT 13

PUR8\_HELPJ

ID PUR8\_HELPJ STANDARD; PRT; 440 AA.

AC Q92KA2;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE ADENYLOSUCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCINASE) (ASL).

DE PURB OR JHP1039.

GN Helicobacter pylori J99 (Campylobacter pylori J99).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI\_TaxID=85963;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,

RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,

RA Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human

RT gastric pathogen Helicobacter pylori.";

RL Nature 397:176-180(1999).

CC -1- CATALYTIC ACTIVITY: 1-(5-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)

CC -5-AMINOIMIDAZOLE + FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-

CC IMIDAZOLECARBOXAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP +

CC FUMARATE + AMP).

CC -1- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.

CC -1- SIMILARITY: BELONGS TO THE LYASE 1 FAMILY. ADENYLOSUCINATE LYASE

CC SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL: AE001531; AAD06609.1; .

DR InterPro: IPR000362; .

DR InterPro: IPR003031; .

DR Pfam: PF00206; lyase\_1; 1.

DR PRINTS: PR00145; DCRYSTALLIN.

DR PRINTS: PR00149; FUMARATE\_LYASE.

DR PROSITE: PS00163; FUMARATE\_LYASES; 1.

KW Purine biosynthesis; Lyase.

FT ACT\_SITE 68 68 ACID (BY SIMILARITY).

FT ACT\_SITE 141 141 BASE (BY SIMILARITY).

```
SQ SEQUENCE 440 AA; 49867 MW; C67D3C02AEFA2EFE CRC64;

Query Match 41.3%; Score 43; DB 1; Length 440;
Best Local Similarity 61.5%; Pred. No. 7.9;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 NFPHGMLDLEETA 15
DB 186 NFAHAPLELEELA 198

RESULT 14
ID PUR8_HELPY STANDARD; PRT; 440 AA.
AC P56468;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENYLOSUCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCINASE) (ASL).
GN PURB OR HP1112.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleschmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Kap P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- CATALYTIC ACTIVITY: 1-(5-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)
CC -5-AMINOMIDAZOLE = FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-
CC IMIDAZOLECARBOXAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP =
CC FUMARATE + AMP)
CC -!- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE LYASE I FAMILY. ADENYLOSUCINATE LYASE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000617; AAD08156.1; -.
CC TIGR; HP1112; -.
CC InterPro; IPR000362; -.
CC Pfam; PF00206; Lyase_1; 1.
CC PRINTS; PR00149; FUMRATLYASE.
CC PROSITE; PS00163; FUMARATE_LYASES; 1.
CC Purine biosynthesis; Lyase.
FT ACT_SITE 68 68 ACID (BY SIMILARITY).
FT ACT_SITE 141 141 BASE (BY SIMILARITY).
SQ SEQUENCE 440 AA; 50028 MW; 59587A4517404A46 CRC64;

Query Match 41.3%; Score 43; DB 1; Length 440;
Best Local Similarity 61.5%; Pred. No. 7.9;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

SQ SEQUENCE 440 AA; 49867 MW; C67D3C02AEFA2EFE CRC64;

Query Match 41.3%; Score 42; DB 1; Length 214;
Best Local Similarity 63.6%; Pred. No. 5.1;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 FPHGMLDLEEI 14
DB 146 FPQGLLDFOEI 156

Search completed: April 24, 2001, 16:34:05
Job time: 561 sec
```

```
OY 3 NFPHGMLDLEETA 15
DB 186 NFAHAPLELEELA 198

RESULT 15
ID GLP1_PHANI STANDARD; PRT; 214 AA.
AC P45853;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GERMIN-LIKE PROTEIN PRECURSOR.
GN GLP.
OS Pharbitis nil (Violet) (Japanese morning glory).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=35883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHOISY; TISSUE=Cotyledon;
RX MEDLINE=97043425; PubMed=8888623;
RA Ono M., Sage-Ono K., Inoue M., Kamada H., Harada H.;
RT "Transient increase in the level of mRNA for a germin-like protein in
RT leaves of the short-day plant Pharbitis nil during the photoperiodic
RT induction of flowering."
RL Plant Cell Physiol. 37:855-861(1996).
CC -!- SIMILARITY: BELONGS TO THE GERMIN (OXALATE OXIDASE) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D45425; BAA08266.1; -.
CC InterPro; IPR001929; -.
CC Pfam; PF01072; Germin; 1.
CC PRINTS; PR00325; GERMIN.
CC PROSITE; PS00725; GERMIN; 1.
CC Cell wall; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 214 GERMIN-LIKE PROTEIN.
SQ SEQUENCE 214 AA; 22608 MW; 2BF187AA8CCAC4AC CRC64;

Query Match 40.4%; Score 42; DB 1; Length 214;
Best Local Similarity 63.6%; Pred. No. 5.1;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 FPHGMLDLEEI 14
DB 146 FPQGLLDFOEI 156

Search completed: April 24, 2001, 16:34:05
Job time: 561 sec
```



**THIS PAGE BLANK (uspto)**

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:33:13 ; Search time 113.97 Seconds  
(without alignments)  
20.568 Million cell updates/sec

Title: US-09-171-432A-44  
Perfect score: 104  
Sequence: 1 KVNFPFGMLDLEETAAANSKD 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	1124	14 Q84780	Q84780 hepatitis a
2	104	100.0	1161	14 Q05794	Q05794 hepatitis a
3	104	100.0	2216	14 Q9WMA2	Q9WMA2 hepatitis a
4	104	100.0	2218	14 Q67824	Q67824 hepatitis a
5	104	100.0	2227	14 Q67825	Q67825 hepatitis a
6	104	100.0	2227	14 Q67826	Q67826 hepatitis a
7	104	100.0	2227	14 Q9WMA4	Q9WMA4 hepatitis a
8	104	100.0	2227	14 Q9WMA3	Q9WMA3 hepatitis a
9	104	100.0	2227	14 Q9WMA1	Q9WMA1 hepatitis a
10	104	100.0	2227	14 Q9WMA0	Q9WMA0 hepatitis a
11	104	100.0	2227	14 Q9IFH5	Q9IFH5 hepatitis a
12	97	93.3	184	14 Q87092	Q87092 simian hepa
13	95	91.3	2218	14 Q67817	Q67817 hepatitis a
14	94	90.4	2227	14 Q9WMA9	Q9WMA9 hepatitis a
15	47	45.2	404	11 Q92218	Q92218 mus musculu
16	47	45.2	1079	5 Q17566	Q17566 caenorhabdi
17	46	44.2	39	5 Q9NC67	Q9NC67 dictyosteli
18	46	44.2	259	2 Q9KPL7	Q9KPL7 vibrio chol
19	46	44.2	677	2 Q25188	Q25188 helicobacte

20 45 43.3 195 4 Q9Y436 Q9Y436 homo sapien  
21 45 43.3 383 10 Q9SYH3 Q9SYH3 arabisdopsis  
22 45 43.3 852 1 Q9YDL5 Q9YDL5 aeropyrum p  
23 45 43.3 935 5 Q62033 Q62033 caenorhabdi  
24 45 43.3 1213 5 Q9XTC7 Q9XTC7 caenorhabdi  
25 44 42.3 154 3 Q9USF0 Q9USF0 schizosacch  
26 44 42.3 223 3 Q13832 Q13832 schizosacch  
27 44 42.3 404 4 Q9S195 Q9S195 homo sapien  
28 43.5 41.8 373 2 P96619 P96619 bacillus su  
29 43 41.3 140 2 Q9PJJ1 Q9PJJ1 chlamydia m  
30 43 41.3 141 2 Q84554 Q84554 chlamydia t  
31 43 41.3 309 1 Q58709 Q58709 pyrococcus  
32 43 41.3 327 10 Q64643 Q64643 arabisdopsis  
33 43 41.3 357 13 Q57429 Q57429 gallus gall  
34 43 41.3 391 13 Q9YI36 Q9YI36 columba liv  
35 43 41.3 465 5 Q9VHJ8 Q9VHJ8 drosophila  
36 43 41.3 465 13 Q9YH21 Q9YH21 gallus gall  
37 43 41.3 506 2 Q50551 Q50551 thermus aqu  
38 43 41.3 587 13 Q9YH20 Q9YH20 gallus gall  
39 43 41.3 917 5 Q45471 Q45471 caenorhabdi  
40 42 40.4 218 5 Q15936 Q15936 aedes aegypt  
41 42 40.4 222 2 Q9KV88 Q9KV88 vibrio chol  
42 42 40.4 229 2 P73268 P73268 synecocyst  
43 42 40.4 309 5 Q76432 Q76432 caenorhabdi  
44 42 40.4 342 2 Q9X041 Q9X041 thermotoga  
45 42 40.4 384 2 Q9KPM7 Q9KPM7 vibrio chol

## ALIGNMENTS

RESULT 1

Q84780 ID Q84780 PRELIMINARY: PRT: 1124 AA.  
AC Q84780:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS)  
DE (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ovchinnikov Y.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,  
RA Rokhlina T.O., Chizhikov V.E., Petrov N.A., Prikhod'ko G.G.,  
RA Blinov V.M., Vasilenko S.K., Sandakhchiev L.S., Kusov Y.Y.,  
RA Grabko V.I., Fleer G.P., Balyan M.S., Drozdov S.G.;  
RL Dokl. Biochem. 285:379-383(1986).  
DR EMBL: X04200; CAA27797.1; -;  
DR EMBL: AL1312; CAA00953.1; -;  
KW Nonstructural protein.  
FT NON\_TER 1  
FT NON\_TER 1124 1124  
SQ SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

Query Match 100.0%; Score 104; DB 14; Length 1124;  
Best Local Similarity 100.0%; Pred. No. 7.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEETAAANSKD 20

Db 879 KVNFPFGMLDLEETAAANSKD 898

RESULT 2

Q05794 ID Q05794 PRELIMINARY: PRT: 1161 AA.  
AC Q05794: Q67801; Q67802; Q67803; Q67804; Q67805; Q67806;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)

01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
P3D (EC 2.7.7.46)) (FRAGMENT).  
Hepatitis A virus.  
Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
Hepatovirus.  
NCBI\_TaxID=12092;  
[1]  
SEQUENCE FROM N.A.  
RA Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilenko S.K.,  
Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,  
Balayan M.S.;  
RL Mol. Gen. Microbiol. Virol. 6:129-133(1987).  
CC -1- SUBUNIT: THE VIRUS CASID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
DR EMBL: X15464; CAA33492.1; -;  
DR INTERPRO: IPR000408; -;  
DR PROSITE: PS00626; RCL\_2; UNKNOWN\_1  
KW Polyprotein; Coat protein; Core protein; RNA-directed RNA polymerase;  
KW Hydroxylase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 >1161 CORE PROTEIN P2C.  
FT NON\_TER 1161 1161  
SQ SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;

Query Match 100.0%; Score 104; DB 14; Length 1161;  
Best Local Similarity 100.0%; Pred. No. 7.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDLEETAAANSKD 20  
|||||  
DB 916 KVNFPHGMLDLEETAAANSKD 935

RESULT 3  
Q9WMA2 PRELIMINARY; PRT; 2216 AA.  
AC Q9WMA2;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AH3;  
RA Fujiwara K.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB020566; BAA35104.1; -;  
DR INTERPRO: IPR000408; -;  
DR INTERPRO: IPR000605; -;  
DR INTERPRO: IPR001205; -;  
DR INTERPRO: IPR001643; -;  
DR PFAM: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM: PF00910; RNA\_helicase; 1.  
DR PRINTS: PR00518; CALICIVIRUS.  
DR PROSITE: PS00626; RCL\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2216 AA; 250209 MW; 1A9D993FEC21FBE82 CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2216;  
Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDLEETAAANSKD 20  
|||||  
DB 922 KVNFPHGMLDLEETAAANSKD 941

RESULT 4  
Q67824 PRELIMINARY; PRT; 2218 AA.  
AC Q67824;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE GBM/FRHK RNA.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GBM;  
RX MEDLINE-94076453; PubMed-8254770;  
RA Graff J., Normann A., Feinstone S.M., Flehmig B.;  
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
RT with two cell culture-adapted variants.";  
RL J. Virol. 68:548-554(1994).  
DR EMBL: X75214; CAA53024.1; -;  
DR INTERPRO: IPR000408; -;  
DR INTERPRO: IPR000605; -;  
DR INTERPRO: IPR001205; -;  
DR PFAM: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM: PF00910; RNA\_helicase; 1.  
DR PROSITE: PS00626; RCL\_2; UNKNOWN\_1.  
FT CHAIN 1 785 P1 STRUCTURAL PROTEINS.  
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEINS.  
FT CHAIN 1417 2218 P3 NONSTRUCTURAL PROTEINS.  
SQ SEQUENCE 2218 AA; 250502 MW; CA72DF0922104C0E CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2218;  
Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KVNFPHGMLDLEETAAANSKD 20  
|||||  
DB 916 KVNFPHGMLDLEETAAANSKD 935

RESULT 5  
Q67825 PRELIMINARY; PRT; 2227 AA.  
AC Q67825;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE GBM/WT RNA.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GBM;  
RX MEDLINE-94076453; PubMed-8254770;  
RA Graff J., Normann A., Feinstone S.M., Flehmig B.;  
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
RT with two cell culture-adapted variants.";  
RL J. Virol. 68:548-554(1994).  
DR EMBL: X75215; CAA53025.1; -;

```
DR INTERPRO: IPR000408; -
DR INTERPRO: IPR000605; -
DR INTERPRO: IPR001205; -
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00910; RNA_helicase; 1.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.
FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 2227 AA; 251563 MW; 4C4D790352F936B4 CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEEIAANSKD 20
Db 922 KVNFPFGMLDLEEIAANSKD 941

RESULT 6
O67826 PRELIMINARY; PRT: 2227 AA.
AC Q67826;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE GBM/HFS RNA.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH2;
RX MEDLINE=94076453; PubMed=8254770;
RA Graff J., Normann A., Feinstein S.M., Flehmig B.;
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison
RT with two cell culture-adapted variants.";
RL J. Virol. 68:548-554 (1994).
DR EMBL: X75216; CAA53026.1; -
DR INTERPRO: IPR000408; -
DR INTERPRO: IPR000605; -
DR INTERPRO: IPR001205; -
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00910; RNA_helicase; 1.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.
FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 2227 AA; 251496 MW; 488CB7C962319457 CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEEIAANSKD 20
Db 922 KVNFPFGMLDLEEIAANSKD 941

RESULT 7
Q9WMA4 PRELIMINARY; PRT: 2227 AA.
AC Q9WMA4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE POLYPROTEIN
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
```

```
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH1;
RA Fujiwara K.;
RT "Hepatitis A virus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBDJ databases.
DR EMBL: AB020564; BAA35102.1; -
DR INTERPRO: IPR000408; -
DR INTERPRO: IPR000605; -
DR INTERPRO: IPR001205; -
DR INTERPRO: IPR001643; -
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00910; RNA_helicase; 1.
DR PRINTS: PR00918; CALICVIRUSN.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251304 MW; 0DEF6D2AEC29C0CE CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEEIAANSKD 20
Db 922 KVNFPFGMLDLEEIAANSKD 941

RESULT 8
Q9WMA3 PRELIMINARY; PRT: 2227 AA.
AC Q9WMA3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH2;
RA Fujiwara K.;
RT "Hepatitis A virus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBDJ databases.
DR EMBL: AB020565; BAA35103.1; -
DR INTERPRO: IPR000408; -
DR INTERPRO: IPR000605; -
DR INTERPRO: IPR001205; -
DR INTERPRO: IPR001643; -
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00910; RNA_helicase; 1.
DR PRINTS: PR00918; CALICVIRUSN.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251440 MW; E04F846CEC7F50FD CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2227;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPFGMLDLEEIAANSKD 20
Db 922 KVNFPFGMLDLEEIAANSKD 941

RESULT 9
Q9WMA1 PRELIMINARY; PRT: 2227 AA.
ID Q9WMA1
```

Q9WMA1;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FH1;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020567; BAA35105.1; -;  
DR INTERPRO; IPR000408; -;  
DR INTERPRO; IPR000605; -;  
DR INTERPRO; IPR001205; -;  
DR INTERPRO; IPR001643; -;  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSN.  
DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2227 AA; 251415 MW; F92C8E2323FC5621 CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDLEETAAANSKD 20  
|||||  
DB 922 KVNFPHGMLDLEETAAANSKD 941

RESULT 10  
Q9WMAO  
ID Q9WMAO PRELIMINARY; PRT; 2227 AA.  
AC Q9WMAO;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FH2;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020568; BAA35106.1; -;  
DR INTERPRO; IPR000408; -;  
DR INTERPRO; IPR000605; -;  
DR INTERPRO; IPR001205; -;  
DR INTERPRO; IPR001643; -;  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSN.  
DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2227 AA; 251118 MW; 53B86B4432127E9B CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDLEETAAANSKD 20

DB 922 KVNFPHGMLDLEETAAANSKD 941  
|||||

RESULT 11  
Q9IFH5  
ID Q9IFH5 PRELIMINARY; PRT; 2227 AA.  
AC Q9IFH5;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE POLYPROTEIN PRECURSOR.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HAF-203;  
RA Baptista M.L., Silva M., de Lima M.A., Yoshida C.F., Gaspar A.M.,  
RA Pires Lopes M.Q., Galler R.;  
RT "Nucleotide sequence of the HAF-203 hepatitis A virus strain isolated  
in Brazil and expression of the VPI gene in a bacterial system.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF268396; AAF80114.1; -;  
SQ SEQUENCE 2227 AA; 251432 MW; 81913AECC8A04200 CRC64;

Query Match 100.0%; Score 104; DB 14; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDLEETAAANSKD 20  
|||||  
DB 922 KVNFPHGMLDLEETAAANSKD 941

RESULT 12  
Q87092  
ID Q87092 PRELIMINARY; PRT; 184 AA.  
AC Q87092;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE SEGMENT 2A-ENCODED PROTEIN (FRAGMENT).  
OS Simian hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90012354; PubMed=2552172;  
RA Brown E.A., Jansen R.W., Lemon S.M.;  
RT "Characterization of a simian hepatitis A virus (HAV): antigenic and  
genetic comparison with human HAV.";  
RL J. Virol. 63:4932-4937(1989).  
DR EMBL; M34085; AAA47483.1; -;  
DR NON\_TER 184  
FT NON\_TER 184  
SQ SEQUENCE 184 AA; 21322 MW; 9FA851FA964A88A8 CRC64;

Query Match 93.3%; Score 97; DB 14; Length 184;  
Best Local Similarity 90.0%; Pred. No. 1.4e-08;  
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVNFPHGMLDLEETAAANSKD 20  
|||||  
DB 131 KVNFPHGMLDLEETAAANSKD 150

RESULT 13  
Q67817



```
ID Q67817 PRELIMINARY; PRT: 2218 AA.
AC Q67817;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE COMPLETE GENOME.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F.G.;
RX MEDLINE=95381623; PubMed=7653108;
RA Beneduce F., Pisani G., Divizia M., Pana A., Morace G.;
RT "Complete nucleotide sequence of a cytopathic hepatitis A virus strain
isolated in Italy."
RL Virus Res. 36:299-309(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F.G.;
RA Morace G.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: X83302; CAA58281.1;
DR INTERPRO: IPR000408;
DR INTERPRO: IPR001205;
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00910; RNA_helicase; 1.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 2218 AA; 250476 MW; 813B21D3E4E533CA CRC64;

Query Match 91.3%; Score 95; DB 14; Length 2218;
Best Local Similarity 95.0%; Pred. No. 5.4e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KVNPHGMLDLEIAANSKD 20
| | | | | | | | | | | | | | | | | | | | | |
DB 916 KVNPHGMLVLEIAANSKD 935

RESULT 14
Q9WM99 PRELIMINARY; PRT: 2227 AA.
AC Q9WM99;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FH3;
RA Fujiwara K.;
RT "Hepatitis A virus."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB020569; BAA35107.1;
DR INTERPRO: IPR000408;
DR INTERPRO: IPR000605;
DR INTERPRO: IPR001205;
DR INTERPRO: IPR001643;
DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
DR PFAM: PF00910; RNA_helicase; 1.
DR PRINTS: PR00918; CALICIVIRUSNS.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251387 MW; 817640187672A23D CRC64;
```

```
Query Match 90.4%; Score 94; DB 14; Length 2227;
Best Local Similarity 90.0%; Pred. No. 8e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KVNPHGMLDLEIAANSKD 20
| | | | | | | | | | | | | | | | | | | | | |
DB 922 KVNPHGMLDLEIAANSKD 941

RESULT 15
Q9Z218 PRELIMINARY; PRT: 404 AA.
AC Q9Z218;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE GTP-SPECIFIC SUCCINYL-COA SYNTHETASE BETA SUBUNIT (FRAGMENT).
GN SUGLG2 OR SCS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=98438536; PubMed=9765291;
RA Johnson J.D., Mehur J.G., Tews K., Milavetz B.I., Lambeth D.O.;
RT "Genetic evidence for the expression of ATP- and GTP-specific
succinyl-CoA synthetases in multicellular eucaryotes."
RL J. Biol. Chem. 273:27580-27586(1998).
CC -|- SIMILARITY: TO OTHER ALPHA SUBUNITS OF SUCCINYL-COA SYNTHETASE, OF
MALATE-COA LIGASE AND TO ATP CITRATE-LYASE.
DR EMBL: AF058956; AAC64399.1;
DR HSP: P07460; 1SCU.
DR MGD: MGI:1306824; Suc1g2.
DR INTERPRO: IPR000303;
DR INTERPRO: IPR003135;
DR PFAM: PF00549; ligase-CoA; 1.
DR PFAM: PF02222; ATP-grasp; 1.
DR PROSITE: PS01217; SUCCINYL_COA_LIG_3; 1.
KW Ligase.
FT NON_TER
SQ SEQUENCE 404 AA; 43857 MW; 511757A75883E8C4 CRC64;
```

```
Query Match 45.2%; Score 47; DB 11; Length 404;
Best Local Similarity 47.4%; Pred. No. 9.7;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 VNFPHGMLDLEIAANSKD 20
| | | | | | | | | | | | | | | | | | | | | |
DB 140 VGSPQGRSDIEEVAASPE 158
```

Search completed: April 24, 2001, 16:33:17  
Job time: 578 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:23:31 ; Search time 121.9 Seconds  
(without alignments)  
9.379 Million cell updates/sec

Title: US-09-171-432A-45

Perfect score: 101

Sequence: 1 DLEEIAANSKDFNMSETDL 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : - A\_Geneseq\_0401.\*

1:	/SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	20	18 W42928	Immunogenic Hepati
2	101	100.0	2227	7 P60066	Sequence of viral
3	101	100.0	2227	11 R05657	Attenuated hepatit
4	101	100.0	2227	18 W34074	Hepatitis A virus
5	101	100.0	2227	21 B18607	Amino acid sequenc
6	101	100.0	2227	21 B18608	Amino acid sequenc
7	101	100.0	2227	21 B18609	Amino acid sequenc
8	53	52.5	20	18 W42927	Immunogenic Hepati
9	43	42.6	152	18 W55452	H. pylori ORF hp6e
10	43	42.6	157	18 W55327	H. pylori ORF hp2p
11	43	42.6	350	20 V27183	(S)-3'-hydroxy-N-m

12	43	42.6	536	21 G48090	Arabidopsis thalia
13	43	42.6	1201	20 W90345	Drosophila sp. Cos
14	42	41.6	177	21 G48507	Arabidopsis thalia
15	42	41.6	226	21 G48506	Arabidopsis thalia
16	42	41.6	265	21 G48505	Arabidopsis thalia
17	42	41.6	365	18 W13493	Pentaerythritol te
18	42	41.6	365	20 Y29461	Enterobacter cloac
19	42	41.6	538	21 Y81713	Streptococcus pneu
20	42	41.6	1837	21 Y85564	Human homologue of
21	41	40.6	163	21 G33763	Arabidopsis thalia
22	41	40.6	166	21 G33762	Arabidopsis thalia
23	41	40.6	177	21 G33761	Arabidopsis thalia
24	41	40.6	178	21 G54807	Arabidopsis thalia
25	41	40.6	289	21 G16959	Arabidopsis thalia
26	41	40.6	289	21 G28545	Arabidopsis thalia
27	41	40.6	292	21 G16958	Arabidopsis thalia
28	41	40.6	292	21 G28544	Arabidopsis thalia
29	41	40.6	376	22 B59176	Thrombin inhibitor
30	41	40.6	376	22 B11125	Human thrombin inh
31	41	40.6	459	21 B56862	Human prostatic can
32	41	40.6	528	21 Y78933	Human UDP-glucuron
33	40.5	40.1	191	20 Y35677	Chlamydia pneumoni
34	40.5	40.1	386	21 Y74678	Neisseria meningit
35	40.5	40.1	386	21 Y74679	Neisseria meningit
36	40.5	40.1	390	21 Y74677	Neisseria gonorrh
37	40.5	40.1	512	21 Y74680	Neisseria gonorrh
38	40.5	40.1	512	21 Y74681	Neisseria meningit
39	40.5	40.1	512	21 Y74682	Neisseria meningit
40	40	39.6	256	21 B27260	Caenorhabditis ele
41	40	39.6	284	21 G16499	Arabidopsis thalia
42	40	39.6	284	21 G39487	Arabidopsis thalia
43	40	39.6	308	21 G39486	Arabidopsis thalia
44	40	39.6	309	21 G16498	Arabidopsis thalia
45	40	39.6	436	21 B42040	Human ORF1804

#### ALIGNMENTS

RESULT 1

W42928

ID W42928 standard; peptide; 20 AA.

XX

AC W42928;

DT 28-APR-1998 (first entry)

XX

DE Immunogenic Hepatitis A virus peptide YK-1328.

XX

KW Immunogenic peptide; immunogenic epitope; P2A protein;

KW Immune response; antibody.

XX

OS Synthetic.

OS Hepatitis A virus.

XX

XX WO9740147-A1.

XX

PD 30-OCT-1997.

XX

PF 18-APR-1997; 97WO-US06891.

XX

PR 19-APR-1996; 96US-0015644.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX

PI Fields HA, Khudyakov YE;

XX

DR WPI; 1997-535831/49.

XX

PT Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an

PT immune response to HAV in a mammal or to detect the presence of

PT antibodies against HAV in a mammal

XX



PR 12-JUL-1988; 88US-0217824.  
 PR 12-JUL-1988; 88US-0652967.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN.  
 XX  
 PI Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;  
 PI Daemer RJ, Gust ID;  
 XX  
 DR WPI; 1990-075557/10.  
 DR N-PSDB; Q03512.  
 XX  
 PT Vaccine against hepatitis A virus infection - comprises novel  
 PT attenuated hepatitis A virus strain.  
 XX  
 PS Claim 1; Fig 1; 18pp; English.  
 XX  
 CC The attenuated HAV is useful for inducing protective immunity against  
 CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
 CC several nucleotide changes distributed throughout the genome, is  
 CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
 CC suitable for use as an HAV vaccine. It is noted that not all the changes  
 CC are necessary for attenuation and use as a vaccine.  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 101; DB 11; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 9e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLEEIAANSKDFPNMSETDL 20  
 DB 931 dleeiaanskdfpnmsetdl 950  
 RESULT 4  
 W34074  
 ID W34074 standard; Protein; 2227 AA.  
 XX  
 AC W34074;  
 XX  
 DT 27-APR-1998 (first entry)  
 XX  
 DE Hepatitis A virus HM-175 protein sequence.  
 XX  
 KW HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;  
 KW infection; vaccine.  
 XX  
 OS Hepatitis A virus HM-175.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..23  
 FT /label= VP4  
 FT Protein 24..245  
 FT /label= VP2  
 FT Protein 246..491  
 FT /label= VP3  
 FT Protein 492..791  
 FT /label= VP1  
 FT Protein 792..980  
 FT /label= 2A  
 FT Protein 981..1087  
 FT /label= 2B  
 FT Protein 1088..1422  
 FT /label= 2C  
 FT Protein 1423..1496  
 FT /label= 3A  
 FT Protein 1497..1519  
 FT /label= 3B  
 FT Protein 1520..1738  
 FT /label= 3C  
 FT Protein 1739..2227  
 FT /label= 3D

XX WO9740166-A2.  
 PN 30-OCT-1997.  
 XX  
 PD 18-APR-1997; 97WO-US06506.  
 PF 19-APR-1996; 96US-0015642.  
 XX  
 PR (USSH ) US SEC DEPT HEALTH.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Emerson SU, Purcell RH, Raychaudhuri G;  
 XX  
 DR WPI; 1997-535850/49.  
 DR N-PSDB; T93023.  
 XX  
 PT Human attenuated HAV genome containing simian HAV 2C gene - useful  
 PT as vaccines against HAV infection  
 XX  
 PS Disclosure; Fig 13A-D; 66pp; English.  
 XX  
 CC This protein sequence is encoded by the human hepatitis A virus  
 CC (HAV) HM-175 wild-type genome (see T93023). Attenuated strain  
 CC HAV/7 is obtained by passage of HM-175 in African Green Monkey  
 CC kidney cells. A claimed DNA construct (1) comprises a genome of  
 CC HAV, where the genome is a human attenuated HAV genome in which a  
 CC region of the 2C gene has been replaced by a corresponding region  
 CC from a 2C gene of a simian AGM-27 HAV genome (see T93024). The  
 CC region of the 2C gene from AGM-27 contained in the construct  
 CC preferably encodes amino acids 120-328 of the 2C protein, amino  
 CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA  
 CC transcript of (1); (2) a cell transfected with (1) or the RNA  
 CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the  
 CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or  
 CC its RNA transcript, can be used as a vaccine for preventing HAV in  
 CC a mammal. (1) or the RNA transcript can also be used to stimulate  
 CC the production of protective antibodies in the mammal.  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 101; DB 18; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 9e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLEEIAANSKDFPNMSETDL 20  
 DB 931 dleeiaanskdfpnmsetdl 950  
 RESULT 5  
 B18607  
 ID B18607 standard; Protein; 2227 AA.  
 XX  
 AC B18607;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.  
 XX  
 KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN US6113912-A.  
 XX  
 PD 05-SEP-2000.  
 XX  
 PF 07-JUN-1995; 95US-0475886.  
 XX  
 PR 18-SEP-1992; 92US-0947338.  
 PR 17-SEP-1993; 93WO-US08610.

```

PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
WPI: 2000-586464/55.
DR N-PSDB; A75476.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Fig 6A-K; 72pp; English.
XX
CC The present sequence is derived from a wild type hepatitis A virus
CC (HAV) strain HM-174. The sequence is modified to produce HAV which
CC are adapted to growth in the human fibroblast-like cell line MRC-5.
CC The HAV is able to propagate in MRC-5 cells and retain appropriate
CC attenuation. It is useful as a live vaccine for prophylaxis of
CC hepatitis A in humans and other primates.
XX
SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 dleeiaanskdfpnmsetdl 950

RESULT 7
B18609
ID B18609 standard; Protein; 2227 AA.
XX
AC B18609;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of live attenuated Hepatitis A virus 4380.
XX
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KW HAV 4380.
XX
OS Hepatitis A virus.
XX
PN US6113912-A.
XX
PD 05-SEP-2000.
XX
PF 07-JUN-1995; 95US-0475886.
XX
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93WO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
WPI: 2000-586464/55.
DR N-PSDB; A75478.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Columns 93-104; 72pp; English.
XX
CC The present sequence is derived from a live attenuated hepatitis A
CC virus (HAV) of the invention, designated HAV 4380. The sequence is
CC produced by modifying wild type HAV strain HM-174. The HAV of the
CC invention are adapted to growth in the human fibroblast-like cell
CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain
CC appropriate attenuation. It is useful as a live vaccine for prophylaxis
CC of hepatitis A in humans and other primates.
XX
SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 dleeiaanskdfpnmsetdl 950

RESULT 6
B18608
ID B18608 standard; Protein; 2227 AA.
XX
AC B18608;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.
XX
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KW P-35 virus.
XX
OS Hepatitis A virus.
XX
PN US6113912-A.
XX
PD 05-SEP-2000.
XX
PF 07-JUN-1995; 95US-0475886.
XX
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93WO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
WPI: 2000-586464/55.
DR N-PSDB; A75477.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Columns 67-78; 72pp; English.
XX
CC The present sequence is derived from passage 35 of a wild type

```

```

CC hepatitis A virus (HAV) strain HM-174. The resulting virus is
CC designated P-35 virus. The sequence is modified to produce HAV which
CC are adapted to growth in the human fibroblast-like cell line MRC-5.
CC The HAV is able to propagate in MRC-5 cells and retain appropriate
CC attenuation. It is useful as a live vaccine for prophylaxis of
CC hepatitis A in humans and other primates.
XX
SQ Sequence 2227 AA;

```

```

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 dleeiaanskdfpnmsetdl 950

```

```

RESULT 7
B18609
ID B18609 standard; Protein; 2227 AA.
XX
AC B18609;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of live attenuated Hepatitis A virus 4380.
XX
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KW HAV 4380.
XX
OS Hepatitis A virus.
XX
PN US6113912-A.
XX
PD 05-SEP-2000.
XX
PF 07-JUN-1995; 95US-0475886.
XX
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93WO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
WPI: 2000-586464/55.
DR N-PSDB; A75478.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Columns 93-104; 72pp; English.
XX
CC The present sequence is derived from a live attenuated hepatitis A
CC virus (HAV) of the invention, designated HAV 4380. The sequence is
CC produced by modifying wild type HAV strain HM-174. The HAV of the
CC invention are adapted to growth in the human fibroblast-like cell
CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain
CC appropriate attenuation. It is useful as a live vaccine for prophylaxis
CC of hepatitis A in humans and other primates.
XX
SQ Sequence 2227 AA;

```

```

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 dleeiaanskdfpnmsetdl 950

```

```

RESULT 6
B18608
ID B18608 standard; Protein; 2227 AA.
XX
AC B18608;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.
XX
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KW P-35 virus.
XX
OS Hepatitis A virus.
XX
PN US6113912-A.
XX
PD 05-SEP-2000.
XX
PF 07-JUN-1995; 95US-0475886.
XX
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93WO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
WPI: 2000-586464/55.
DR N-PSDB; A75477.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Columns 67-78; 72pp; English.
XX
CC The present sequence is derived from passage 35 of a wild type

```

```

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 dleeiaanskdfpnmsetdl 950

```

```

RESULT 7
B18609
ID B18609 standard; Protein; 2227 AA.
XX
AC B18609;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of live attenuated Hepatitis A virus 4380.
XX
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KW HAV 4380.
XX
OS Hepatitis A virus.
XX
PN US6113912-A.
XX
PD 05-SEP-2000.
XX
PF 07-JUN-1995; 95US-0475886.
XX
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93WO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
WPI: 2000-586464/55.
DR N-PSDB; A75478.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Columns 93-104; 72pp; English.
XX
CC The present sequence is derived from a live attenuated hepatitis A
CC virus (HAV) of the invention, designated HAV 4380. The sequence is
CC produced by modifying wild type HAV strain HM-174. The HAV of the
CC invention are adapted to growth in the human fibroblast-like cell
CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain
CC appropriate attenuation. It is useful as a live vaccine for prophylaxis
CC of hepatitis A in humans and other primates.
XX
SQ Sequence 2227 AA;

```

```

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 dleeiaanskdfpnmsetdl 950

```

```

RESULT 6
B18608
ID B18608 standard; Protein; 2227 AA.
XX
AC B18608;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.
XX
KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
KW P-35 virus.
XX
OS Hepatitis A virus.
XX
PN US6113912-A.
XX
PD 05-SEP-2000.
XX
PF 07-JUN-1995; 95US-0475886.
XX
PR 18-SEP-1992; 92US-0947338.
PR 17-SEP-1993; 93WO-US08610.
PR 10-MAR-1995; 95US-0397232.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX
WPI: 2000-586464/55.
DR N-PSDB; A75477.
XX
PT Novel live hepatitis A virus adapted to growth in human fibroblast cell
PT line useful as vaccine for protecting humans against hepatitis A virus
PT infection, has modified genome compared to wild type
XX
PS Disclosure; Columns 67-78; 72pp; English.
XX
CC The present sequence is derived from passage 35 of a wild type

```

```

Query Match 100.0%; Score 101; DB 21; Length 2227;
Best Local Similarity 100.0%; Pred. No. 9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 dleeiaanskdfpnmsetdl 950

```

Db 931 dleelaanskdipnmsetdl 950  
|||||

RESULT 8  
W42927 W42927 standard; peptide; 20 AA.  
XX AC W42927;  
XX DT 28-APR-1998 (first entry)  
XX DE Immunogenic Hepatitis A virus peptide YK-1327.  
XX KW Immunogenic peptide; immunogenic epitope; P2A protein;  
KW KW Immune response; antibody.  
XX OS Synthetic.  
XX OS Hepatitis A virus.  
XX PN W09740147-Al.  
XX PD 30-OCT-1997.  
XX PF 18-APR-1997; 97WO-US06891.  
XX PR 19-APR-1996; 96US-0015644.  
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX PI Fields HA, Khudyakov YE;  
XX PS Claim 18; Page 112; 140pp; English.  
XX CC Peptides W42922-30 are immunogenic peptides corresponding to immunogenic  
CC epitopes of the Hepatitis A virus (HAV). The peptides are substantially  
CC similar to a portion of the amino acid sequence of the P2A protein of HAV  
CC corresponding to amino acids 792-980. The present peptide is derived  
CC from amino acids 922-941, and has a reactivity of 31.3% with acute sera.  
CC Compositions containing the peptides can be used to induce an immune  
CC response to HAV in a mammal. The peptides can also be used to detect the  
CC presence of antibodies against HAV in mammalian serum. The peptides can  
CC also be used to make an antibody against HAV by administering the peptide  
CC to a mammal.

XX Sequence 20 AA;  
Query Match 52.5%; Score 53; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.031;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKD 11  
|||||  
Db 10 dleelaanskd 20

RESULT 9  
W55492 W55492 standard; Protein; 152 AA.  
XX ID W55492;  
XX AC W55492;  
XX DT 30-JUN-1998 (first entry)  
XX DE H: pylori ORF hp6el0967\_24889750\_f2\_7 secreted protein.  
XX

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
XX bacterium.  
XX OS Helicobacter pylori.  
XX PN W09737044-Al.  
XX PD 09-OCT-1997.  
XX PF 27-MAR-1997; 97WO-US05223.  
XX PR 06-DEC-1996; 96US-0761318.  
XX PR 29-MAR-1996; 96US-0625811.  
XX PR 02-APR-1996; 96US-0758731.  
XX PR 25-OCT-1996; 96US-0736905.  
XX PR 28-OCT-1996; 96US-0738859.  
XX PA (ASTR ) ASTRA AB.  
XX PI Alm RA, Smith D;  
XX DR WPI; 1997-503122/46.  
XX DR N-PSDB; V24901.  
XX PT Helicobacter pylori nucleic acid sequences and encoded  
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection  
XX PS Claims 14,94; Page 698; 1145pp; English.  
XX CC This sequence is a H. pylori secreted protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The DNA and probes derived from it may be used for the  
CC identification of H. pylori in a sample, and the diagnosis of  
CC H. pylori infection. Nucleic acid sequences complementary to the  
CC DNA act as antisense sequences, and can be used to prevent the  
CC translation of H. pylori mRNA. Antibodies against the protein can  
CC be used in immunoassays to evaluate the abundance and distribution  
CC of H. pylori-specific antigens. The genomic sequence of H. pylori  
CC (ATCC 55679) was determined from overlapping contigs generated  
CC by mechanically shearing the bacterial DNA. The sequences were  
CC analysed for ORF of at least 180 nucleotides, and the predicted  
CC coding regions defined by computer evaluation. To identify likely  
CC H. pylori antigens for vaccine development, the amino acid  
CC sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having  
CC identified and determined the sequences of interest, particular  
CC regions can be isolated from H. pylori by PCR amplification for  
CC recombinant polypeptide production, e.g. in E. coli hosts.

XX Sequence 152 AA;

Query Match 42.6%; Score 43; DB 18; Length 152;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 IAANSKDFPNMSETDL 20  
|||  
Db 16 valnakdfsktsdcl 31  
RESULT 10  
W55327 ID W55327 standard; Protein; 157 AA.  
XX AC W55327;  
XX DT 15-JUN-1998 (first entry)

XX H. pylori ORF hp2p10625orf6 protein.  
 DE Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 XX Helicobacter pylori.  
 OS  
 XX  
 XX WO9737044-A1.  
 PN  
 XX  
 XX 09-OCT-1997.  
 PD  
 XX  
 XX 27-MAR-1997; 97WO-US05223.  
 PF  
 XX  
 PR 06-DEC-1996; 96US-0761318.  
 PR 29-MAR-1996; 96US-0625811.  
 PR 02-APR-1996; 96US-0758731.  
 PR 25-OCT-1996; 96US-0736905.  
 PR 28-OCT-1996; 96US-0738859.  
 XX  
 XX (ASTR ) ASTRA AB.  
 PA  
 XX Alm RA, Smith D;  
 PI  
 XX  
 XX WPI: 1997-503122/46.  
 DR N-PSDB; V24736.  
 DR  
 XX Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 PT  
 XX  
 XX Claim 14; Pages 553-554; 1145pp; English.  
 PS  
 XX This sequence is a H. pylori protein of unspecified function.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 XX  
 XX Sequence 157 AA;  
 SQ  
 Query Match 42.6%; Score 43; DB 18; Length 157;  
 Best Local Similarity 50.0%; Pred. No. 15;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 Qy 5 1AANKDFPNMSETDL 20  
 Db 21 valnakdfsktsdedl 36  
 RESULT 11  
 Y27183  
 ID Y27183 standard; Protein; 350 AA.  
 XX  
 XX Y27183;  
 AC  
 XX

DT 20-SEP-1999 (first entry)  
 XX  
 DE (S)-3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase (4'-OMT).  
 XX  
 KW (S)-3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase; 4'-OMT;  
 KW enzyme; secondary metabolite; reticuline; drug.  
 XX  
 XX Coptis japonika.  
 OS  
 XX  
 XX JP11178579-A.  
 PN  
 XX  
 XX 06-JUL-1999.  
 PD  
 XX  
 XX 24-DEC-1997; 97JP-0355320.  
 PF  
 XX  
 XX 24-DEC-1997; 97JP-0355320.  
 PR  
 XX (MITC ) MITSUI PETROCHEM IND CO LTD.  
 PA  
 XX WPI: 1999-437312/37.  
 DR N-PSDB; X89255.  
 DR  
 XX New (S)-3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase - and  
 PT gene encoding it  
 PT  
 XX Claim 1; Fig 12; 28pp; Japanese.  
 PS  
 XX This represents a polypeptide having the enzymatic activity of (S)-3'-  
 CC -hydroxy-N-methylcoclaurine 4'-O-methyltransferase (4'-OMT). A vector  
 CC containing the DNA can be used to transform a microbe for the recombinant  
 CC preparation of the enzyme. The enzyme can be used to prepare a secondary  
 CC metabolite of a plant derived from reticuline which is useful as a raw  
 CC material for drugs.  
 CC  
 XX Sequence 350 AA;  
 SQ  
 Query Match 42.6%; Score 43; DB 20; Length 350;  
 Best Local Similarity 53.3%; Pred. No. 37;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 DLEEIAANSKDFPNM 15  
 Db 219 dlphviansydlpni 233  
 RESULT 12  
 G48090  
 ID G48090 standard; Protein; 536 AA.  
 XX  
 AC G48090;  
 XX  
 XX 18-OCT-2000 (first entry)  
 DT  
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 60693.  
 DE  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 XX Arabidopsis thaliana.  
 OS  
 XX  
 XX EP1033405-A2.  
 PN  
 XX  
 XX 06-SEP-2000.  
 PD  
 XX  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF  
 XX  
 XX 25-FEB-1999; 99US-0121825.  
 PR  
 XX 05-MAR-1999; 99US-0123180.  
 PR  
 XX 09-MAR-1999; 99US-0123548.  
 PR  
 XX 23-MAR-1999; 99US-0125788.  
 PR  
 XX 25-MAR-1999; 99US-0126264.  
 PR



PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

```

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 42.6%; Score 43; DB 21; Length 536;
Best Local Similarity 47.4%; Pred. No. 61;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 DLEETAAANSKDFPNMSETD 19
| | : : | : | | | |
Db 21 dniednsskyendsetd 39

RESULT 13
W90345
ID W90345 standard; Protein; 1201 AA.
XX AC
XX DT
XX 25-MAY-1999 (first entry)
XX DE Drosophila sp. Cos2 protein.
XX
XX Cos2; fruitfly; Costal2; Kinesin-related; hedgehog; modulate; mapping;
KW gene therapy; prophylactic; therapeutic; treatment; cancer; meningioma;
KW basal cell carcinoma; glioma; medulloblastoma; breast carcinoma;
KW hedgehog signalling pathway; infertility treatment; contraception;
KW developmental abnormalities; immunogen.
XX
XX Drosophila sp.
XX
XX Key Location/Qualifiers
FH Region 1..450
FT /note= "N-terminal region"
FT Binding-site 174..181
FT /note= "Putative nucleotide binding motif N1 (P-loop)"
FT Binding-site 237..239
FT /note= "Putative microtubule binding motif"
FT Binding-site 265..269
FT /note= "Putative nucleotide binding motif N2"
FT Binding-site 298..303
FT /note= "Putative nucleotide binding motif N3"
FT Binding-site 340..345
FT /note= "Putative microtubule binding motif"
FT Region 543..990
FT /note= "central region"
FT Region 643..650

```

```

FT Region
FT /note= "heptad repeat"
FT 651..656
FT /note= "heptad repeat"
FT 657..663
FT /note= "heptad repeat"
FT 664..670
FT /note= "heptad repeat"
FT 671..677
FT /note= "heptad repeat"
FT 678..684
FT /note= "heptad repeat"
FT 685..691
FT /note= "heptad repeat"
FT 692..698
FT /note= "heptad repeat"
FT 706..712
FT /note= "heptad repeat"
FT 713..719
FT /note= "heptad repeat"
FT 720..726
FT /note= "heptad repeat"
FT 727..733
FT /note= "heptad repeat"
FT 740..746
FT /note= "heptad repeat"
FT 747..753
FT /note= "heptad repeat"
FT 754..760
FT /note= "heptad repeat"
FT 761..767
FT /note= "heptad repeat"
FT 771..777
FT /note= "heptad repeat"
FT 778..784
FT /note= "heptad repeat"
FT 785..791
FT /note= "heptad repeat"
FT 792..798
FT /note= "heptad repeat"
FT 799..805
FT /note= "heptad repeat"
FT 806..812
FT /note= "heptad repeat"
FT 833..839
FT /note= "heptad repeat"
FT 840..846
FT /note= "heptad repeat"
FT 847..853
FT /note= "heptad repeat"
FT Misc-difference 857
FT /label= unknown
FT Region 870..876
FT /note= "heptad repeat"
FT Region 877..883
FT /note= "heptad repeat"
FT Region 884..890
FT /note= "heptad repeat"
FT Region 905..911
FT /note= "heptad repeat"
FT Region 912..918
FT /note= "heptad repeat"
FT Region 919..925
FT /note= "heptad repeat"
FT Region 926..932
FT /note= "heptad repeat"
FT Region 963..969
FT /note= "heptad repeat"
FT Region 970..976
FT /note= "heptad repeat"
FT Region 977..983
FT /note= "heptad repeat"
FT Region 984..990
FT /note= "heptad repeat"
FT

```

FT Region 1050..1201  
 FT /note= "C-terminal region"  
 XX WO9900403-A1.  
 XX 07-JAN-1999.  
 XX 18-JUN-1998; 98WO-US12820.  
 XX 30-JUN-1997; 97US-0051347.  
 XX (STRD ) UNIV LELAND STANFORD JUNIOR.  
 XX Scott MP, Sisson JC;  
 XX WPI; 1999-095669/08.  
 DR N-PSDB; V72291.  
 PT New costal2 (cos2) gene - used to identify homologous and related  
 PT proteins  
 XX Claim 3; Page 31-34; 42pp: English.  
 XX This sequence represents a Cos2 gene isolated from *Drosophila* sp. The  
 CC Cos2 protein (also known as the Costal2 protein) is a kinesin-related  
 CC protein that accumulates preferentially in cells capable of responding to  
 CC hedgehog. This sequence is used to identify homologous or related genes,  
 CC in producing compositions that modulate the expression or function of the  
 CC encoded protein, for gene therapy, mapping functional regions of the  
 CC protein and in studying associated physiological pathways. In addition,  
 CC modulation of Cos2 gene activity is used for prophylactic and therapeutic  
 CC purposes such as treatment of cancers (e.g. basal cell carcinoma, glioma,  
 CC medulloblastoma, meningioma, breast carcinoma), hedgehog signalling  
 CC pathway function, identification of cell type based on expression,  
 CC infertility treatment, contraception, developmental abnormalities. The  
 CC encoded protein is useful as an immunogen for producing specific  
 CC antibodies, in screening for biologically active agents that modulate  
 CC Cos2 function and that act in the hedgehog signalling pathway. The  
 CC protein can also be used for therapeutic and prophylactic purposes.  
 XX Sequence 1201 AA;  
 QY 4 EIAANSKDFNMSETD 19  
 Db 48 qvpaekdnpsisetd 63  
 Query Match 42.6%; Score 43; DB 20; Length 1201;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 RESULT 14  
 G48507  
 ID G48507 standard; Protein; 177 AA.  
 XX G48507;  
 AC  
 XX 18-OCT-2000 (first entry)  
 DT  
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 61261.  
 DE  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 OS  
 XX EP1033405-A2.  
 PN  
 XX 06-SEP-2000.  
 PD  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF

XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0132863.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.

```
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 17-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154038.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.6%; Score 42; DB 21; Length 177;
Best Local Similarity 36.8%; Pred. No. 25;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 1 DLEELANSKDFPNMSETD 19
|| : || : || :
Db 157 dlaplanahsygpmhvd 175

RESULT 15
G48506
ID G48506 standard; Protein; 226 AA.
XX G48506;
AC G48506;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61260.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
```

PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
Query Match 41.6%; Score 42; DB 21; Length 226;  
Best Local Similarity 36.8%; Pred. NO. 33;  
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 DLEETAANSKDFPNNSETD 19  
|| : || : | | : |  
Db 206 dlapvianahsyppgmhvd 224

Search completed: April 24, 2001, 16:23:33  
Job time: 151 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:35:05 ; Search time 62.1 Seconds  
(without alignments)  
6.187 Million cell updates/sec

Title: US-09-171-432a-45

Perfect score: 101

Sequence: 1 DLEEIAANSKDFPNMSETDL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2.6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2.6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2.6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2.6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2.6/prodata/2/1aa/PCUS\_COMB.pep:\*  
6: /cgn2.6/prodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	2227	4 US-08-475-886-2	Sequence 2, Appli
2	101	100.0	2227	4 US-08-475-886-4	Sequence 4, Appli
3	101	100.0	2227	4 US-08-475-886-6	Sequence 6, Appli
4	101	100.0	2227	4 US-08-397-232-2	Sequence 2, Appli
5	101	100.0	2227	4 US-08-397-232-4	Sequence 4, Appli
6	44	43.6	907	3 US-08-938-830-26	Sequence 26, Appli
7	44	43.6	907	4 US-09-020-222-26	Sequence 26, Appli
8	42	41.6	365	2 US-08-983-352-2	Sequence 2, Appli
9	41	40.6	376	4 US-09-200-965-2	Sequence 2, Appli
10	39	38.6	50	1 US-08-127-351-14	Sequence 14, Appli
11	39	38.6	50	1 US-08-480-367B-14	Sequence 14, Appli
12	39	38.6	50	1 US-08-487-221A-14	Sequence 14, Appli
13	39	38.6	50	1 US-08-480-370-14	Sequence 14, Appli
14	39	38.6	352	4 US-09-286-691-26	Sequence 26, Appli
15	39	38.6	396	1 US-08-430-024-2	Sequence 2, Appli
16	39	38.6	396	1 US-08-782-009-2	Sequence 2, Appli
17	39	38.6	396	3 US-09-017-302-2	Sequence 2, Appli
18	39	38.6	984	1 US-08-242-932-2	Sequence 2, Appli
19	39	38.6	984	1 US-08-714-481-2	Sequence 2, Appli
20	39	38.6	984	5 PCT-US95-06111-2	Sequence 2, Appli
21	38	37.6	99	1 US-08-004-492-9	Sequence 9, Appli
22	38	37.6	99	2 US-08-484-905-60	Sequence 60, Appli
23	38	37.6	99	3 US-08-481-985B-60	Sequence 60, Appli
24	38	37.6	99	4 US-08-370-476-60	Sequence 60, Appli
25	38	37.6	105	1 US-08-081-539-73	Sequence 73, Appli
26	38	37.6	105	1 US-08-466-647-73	Sequence 73, Appli
27	38	37.6	114	4 US-09-347-819-4	Sequence 4, Appli

```

28 38 37.6 162 2 US-08-716-317-10 Sequence 10, Appl
29 38 37.6 163 2 US-08-716-317-15 Sequence 15, Appl
30 38 37.6 164 1 US-08-318-193-60 Sequence 60, Appl
31 38 37.6 172 4 US-08-149-101A-21 Sequence 21, Appl
32 38 37.6 172 5 PCT-US94-12873-21 Sequence 21, Appl
33 38 37.6 174 4 US-08-149-101A-20 Sequence 20, Appl
34 38 37.6 174 5 PCT-US94-12873-20 Sequence 20, Appl
35 38 37.6 181 1 US-08-209-182C-4 Sequence 4, Appl
36 38 37.6 181 5 PCT-US92-05612-4 Sequence 4, Appl
37 38 37.6 182 4 US-08-149-101A-25 Sequence 25, Appl
38 38 37.6 182 5 PCT-US94-12873-25 Sequence 25, Appl
39 38 37.6 183 1 US-08-009-973-1 Sequence 1, Appl
40 38 37.6 184 1 US-08-567-047-2 Sequence 2, Appl
41 38 37.6 184 2 US-08-693-182-2 Sequence 2, Appl
42 38 37.6 184 2 US-08-567-048-2 Sequence 2, Appl
43 38 37.6 184 2 US-09-008-482-2 Sequence 8, Appl
44 38 37.6 184 2 US-08-945-529-8 Sequence 8, Appl
45 38 37.6 184 6 5186931-1 Patent No. 5186931

```

## ALIGNMENTS

```

RESULT 1
US-08-475-886-2
; Sequence 2, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2

```

```

Query Match 100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 DLEEIAANSKDFPNMSETDL 20
Db 931 DLEEIAANSKDFPNMSETDL 950

```

```

RESULT 2
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18

```

```
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETAAANSKDFPNMSETDL 20
   |||||||
DB 931 DLEETAAANSKDFPNMSETDL 950

RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETAAANSKDFPNMSETDL 20
   |||||||
DB 931 DLEETAAANSKDFPNMSETDL 950

RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-2

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETAAANSKDFPNMSETDL 20
   |||||||
DB 931 DLEETAAANSKDFPNMSETDL 950

RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETAAANSKDFPNMSETDL 20
   |||||||
DB 931 DLEETAAANSKDFPNMSETDL 950

RESULT 6
US-08-938-830-26
; Sequence 26, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```



;; SOFTWARE: WinPatIn (Genentech)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/938,830  
;; FILING DATE:  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/798419  
;; FILING DATE: 07-FEB-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dreger, Ginger R.  
;; REGISTRATION NUMBER: 33,055  
;; REFERENCE/DOCKET NUMBER: P1066P1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 650/225-3216  
;; TELEFAX: 650/952-9881  
;; INFORMATION FOR SEQ ID NO: 26:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 907 amino acids  
;; TYPE: Amino Acid  
;; TOPOLOGY: Linear  
US-08-938-830-26

Query Match 43.6%; Score 44; DB 3; Length 907;  
Best Local Similarity 47.1%; Pred. No. 25;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 2 LEEIAANSKDFPNMSET 18  
Db 49 LQELAASSADIPVSGT 65

RESULT 7  
US-09-020-222-26  
; Sequence 26, Application US/09020222  
; Patent No. 6111073  
; GENERAL INFORMATION:  
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage  
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPS)  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/020,222  
; FILING DATE: 06-Feb-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/798419  
; FILING DATE: 02/07/1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: P1066P1  
; TELEPHONE: 650/225-3216  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 907 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
US-09-020-222-26

Query Match 43.6%; Score 44; DB 4; Length 907;  
Best Local Similarity 47.1%; Pred. No. 25;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 2 LEEIAANSKDFPNMSET 18  
Db 49 LQELAASSADIPVSGT 65

RESULT 8  
US-08-983-352-2  
; Sequence 2, Application US/08983352  
; Patent No. 5928859  
; GENERAL INFORMATION:  
; APPLICANT: Nicklin, Stephen  
; APPLICANT: Binks, Peter R.  
; APPLICANT: Bruce, Neil C.  
; APPLICANT: French, Christopher E.  
; TITLE OF INVENTION: DETECTION AND BIODEGRADATION OF  
; TITLE OF INVENTION: EXPLOSIVES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 No. 5928859th Glebe Rd. 8th floor  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201-4741  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/983,352  
; FILING DATE: 02-FEB-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01629  
; FILING DATE: 08-JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9514138.8  
; FILING DATE: 11-JUL-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9607220.2  
; FILING DATE: 04-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford, Arthur R.  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 124-617  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 365 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-983-352-2

Query Match 41.6%; Score 42; DB 2; Length 365;  
Best Local Similarity 47.4%; Pred. No. 18;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 2 LEEIAANSKDFPNMSETDL 20  
Db 258 IEELAKRGIAIYLMSETDL 276

[illegible]

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-480-367B-14

Query Match 38.6%; Score 39; DB 1; Length 50;  
Best Local Similarity 43.8%; Pred. No. 5.2;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 EIAANSKDFPNMSETD 19  
|:|:|:|:|:|:|

DB 15 ELAENNLNPKMAEKD 30

RESULT 12

US-08-487-221A-14  
Sequence 14, Application US/08487221A  
Patent No. 5593656

GENERAL INFORMATION:

APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.

APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT, P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487.221A  
FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/127,351  
FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-487-221A-14

Query Match 38.6%; Score 39; DB 1; Length 50;  
Best Local Similarity 43.8%; Pred. No. 5.2;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 EIAANSKDFPNMSETD 19  
|:|:|:|:|:|:|

DB 15 ELAENNLNPKMAEKD 30

RESULT 13

US-08-480-370-14  
Sequence 14, Application US/08480370  
Patent No. 5609847

GENERAL INFORMATION:

APPLICANT: BELINKA JR, BENJAMIN A.  
APPLICANT: COUGHLIN, DANIEL J.  
APPLICANT: ALVAREZ, VERNON L.

APPLICANT: WOOD, RICHARD  
TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE  
TITLE OF INVENTION: CONSTRUCTS

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT, P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480.370  
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/127,351  
FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Villacorta, Gilberto M.  
REGISTRATION NUMBER: 34,038  
REFERENCE/DOCKET NUMBER: 4980-004-44

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-480-370-14

Query Match 38.6%; Score 39; DB 1; Length 50;  
Best Local Similarity 43.8%; Pred. No. 5.2;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 EIAANSKDFPNMSETD 19  
|:|:|:|:|:|:|

DB 15 ELAENNLNPKMAEKD 30

RESULT 14

US-09-286-691-26  
Sequence 26, Application US/09286691  
Patent No. 6190189

GENERAL INFORMATION:

APPLICANT: Li, Xin-Liang  
APPLICANT: Ljungdahl, Lars G.  
APPLICANT: Chen, Huizhong

TITLE OF INVENTION: Cellulases and Coding Sequences  
FILE REFERENCE: 42-96  
CURRENT APPLICATION NUMBER: US/09/286.691

```

Query Match      38.6%; Score 39; DB 1; Length 396;
Best Local Similarity 50.0%; Pred. NO. 63;
Matches 8; Conservative 5; Mismatches 1; Indels 2; Gaps 1;

QY      4      EIAANSKDFPNMSETD 19
      |::| |::|::|::|

```



**THIS PAGE BLANK (uspto)**

GenCore version 4.5  
Copyright (c).1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:25:59 ; Search time 73.14 Seconds  
(without alignments)  
18.792 Million cell updates/sec

Title: US-09-171-432A-45  
Perfect score: 101  
Sequence: 1 DLEEIAANSKDFPNSETDL 20  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_67.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed; and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	2227	1 GNNYHM	genome polyprotein
2	101	100.0	2227	1 GNNYHR	genome polyprotein
3	101	100.0	2227	1 GNNYMK	genome polyprotein
4	101	100.0	2227	1 GNNYHB	genome polyprotein
5	98	97.0	2230	1 GNNYSA	genome polyprotein
6	45.5	45.0	1621	2 T15264	hypothetical prote
7	44	43.6	740	1 FOLJHD	gag polyprotein - f
8	44	43.6	927	2 T38127	phosphoprotein - f
9	44	43.6	1046	2 S67786	hypothetical prote
10	43.5	43.1	754	1 BABOH	peptide-aspartate
11	43.5	43.1	1451	2 S65571	pattern formation
12	43.5	43.1	5107	2 T29144	partial CDS - Caen
13	43	42.6	152	2 A64610	hypothetical prote
14	43	42.6	152	2 A71904	hypothetical prote
15	43	42.6	231	2 H70407	probable di-trans,
16	43	42.6	454	2 T02100	hypothetical prote
17	43	42.6	598	2 H71336	probable cell divi
18	43	42.6	1072	2 A84112	alkaline amylopul
19	43	42.6	1201	2 T08603	kinesin-related pr
20	42.5	42.1	913	2 T15278	hypothetical prote
21	42.5	42.1	4717	2 T41581	hypothetical coile
22	42	41.6	289	2 T23342	hypothetical prote
23	42	41.6	365	2 C81050	cytochrome c oxida
24	42	41.6	365	2 F81826	probable cytochrom
25	42	41.6	1112	2 T47784	hypothetical prote
26	41.5	41.1	236	2 A81283	probable periplasm
27	41.5	41.1	694	2 S41868	DNA-directed RNA p
28	41.5	41.1	1378	2 A81393	DNA-directed RNA p
29	41	40.6	94	2 D70245	hypothetical prote

30	41	40.6	133	2	A29174	ciostripain (EC 3.
31	41	40.6	139	2	T26224	hypothetical prote
32	41	40.6	230	2	T45754	hypothetical prote
33	41	40.6	295	2	H83642	probable 2-OH-laur
34	41	40.6	346	2	G64182	lipopolysaccharide
35	41	40.6	376	1	A48681	placental thrombin
36	41	40.6	387	2	F72416	alcohol dehydrogen
37	41	40.6	439	2	C32812	repC protein - Agr
38	41	40.6	493	2	T31794	hypothetical prote
39	41	40.6	526	2	S35190	ciostripain (EC 3.
40	41	40.6	528	2	JN0619	glucuronosyltransf
41	41	40.6	548	2	B82486	hemolysin secretio
42	41	40.6	548	2	S15910	hemolysin-releasin
43	41	40.6	657	2	E81119	tail fibre protein
44	41	40.6	657	2	E81901	probable phage tai
45	41	40.6	686	2	C83734	beta-hexosamidase

## ALIGNMENTS

## RESULT 1

GNNYHM genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
B; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with  
A:Reference number: A25981; MUID:87061253  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: core protein 3A #status predicted <C3A>  
F:1497-1519/Product: core protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNSETDL 20  
|||||  
DB 931 DLEEIAANSKDFPNSETDL 950

## RESULT 2

GNNYHR genome polyprotein - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03903  
R:Najarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van N  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985





A:Cross-references: EMBL:X15461  
 C:Superfamily: hepatitis A virus genome polyprotein  
 C:Keywords: coat protein; core protein; polyprotein  
 F:1-163/Product: coat protein 1A #status predicted <C1A>  
 F:28-249/Product: coat protein 1B #status predicted <C1B>  
 F:250-495/Product: coat protein 1C #status predicted <C1C>  
 F:496-795/Product: coat protein 1D #status predicted <C1D>  
 F:796-984/Product: coat protein 2A #status predicted <C2A>  
 F:1092-1426/Product: coat protein 2B #status predicted <C2B>  
 F:1427-1498/Product: coat protein 2C #status predicted <C2C>  
 F:1499-1521/Product: coat protein 3A #status predicted <C3A>  
 F:1522-1741/Product: coat protein 3B #status predicted <C3B>  
 F:1742-2230/Product: coat protein 3D #status predicted <C3D>

Query Match 97.0%; Score 98; DB 1; Length 2230;  
 Best Local Similarity 95.0%; Pred. No. 4.7e-07;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETAANSKDFPNMSETDL 20

Db 935 DLEETAANSKDFPNMSETDL 954

RESULT 6

T15264

hypothetical protein F59E12.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T15264

R:Johnson, D.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid F59E12.

A:Reference number: Z18318

A:Accession: T15264

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1621 <JOH>

A:Cross-references: EMBL:AF003386; NID:g2088833; PID:g2088843; PIDN:AAB54259.1; GSPDB:GN

A:Experimental source: strain Bristol N2; clone F59E12

C:Genetics:

A:Gene: CESP:F59E12.9

A:Map position: 2

A:Introns: 30/3; 55/1; 200/2; 299/2; 327/2; 369/3; 589/3; 860/1; 986/1; 1278/1; 1547/1

Query Match 45.0%; Score 45.5; DB 2; Length 1621;

Best Local Similarity 42.9%; Pred. No. 75;

Matches 9; Conservative 8; Mismatches 1; Indels 3; Gaps 1;

QY 2 LEEIAANSKDF---PNMSETD 19

Db 985 IREIVASAKDFMNDPVSQSD 1005

RESULT 7

FOLJHD

gag polyprotein - squirrel monkey retrovirus SMRV-H

N:Contents: core protein p16; core protein p19; probable core protein p10; probable core

C:Species: squirrel monkey retrovirus SMRV-H

C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999

C:Accession: A31827

R:Oda, T.; Ikeda, S.; Watanabe, S.; Akiyama, K.; Mitsunobu, F.

Virology 167, 468-476, 1988

A:Title: Molecular cloning, complete nucleotide sequence, and gene structure of the prov

A:Reference number: A31827; MUID:89073750

A:Accession: A31827

A:Molecule type: DNA

A:Residues: 1-740 <ODA>

A:Cross-references: GB:M23385; NID:g332626; PIDN:AAA66451.1; PID:g807672

C:Genetics:

A:Gene: gag

C:Superfamily: AIDS-related virus gag polyprotein  
 C:Keywords: core protein; polyprotein  
 F:1-163/Product: core protein p19 #status predicted <CP9>  
 F:164-318/Product: core protein p16 #status predicted <CP6>  
 F:319-648/Product: core protein p35 #status predicted <CP5>  
 F:649-740/Product: core protein p10 #status predicted <CP1>

Query Match 43.6%; Score 44; DB 1; Length 740;

Best Local Similarity 50.0%; Pred. No. 53;

Matches 10; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

QY 1 DLEETAA--NSKDFPNMSET 18

Db 214 DLEETAAQYNNPDWQLTNT 233

RESULT 8

T38127

phosphoprotein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 01-Dec-2000

C:Accession: T38127; A57087; S54119

R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1997

A:Reference number: Z21772

A:Accession: T38127

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-927 <BAD>

A:Cross-references: EMBL:Z95334; PIDN:CAB08599.2; GSPDB:GN000066; SPDB:SPAC20G8.05c

A:Experimental source: strain 972h; cosmid c20G8

R:Fankhauser, C.; Raymond, A.; Cerutti, L.; Utzig, S.; Hofmann, K.; Simanis, V.

Cell 82, 435-444, 1995

A:Title: The Schizosaccharomyces pombe cdc15 gene is a key element in the reorganizat

A:Reference number: A57087; MUID:95360987

A:Accession: A57087

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 'MLTKSLQ', 28-927 <FAN>

A:Cross-references: GB:X86179

R:Fankhauser, C.; Raymond, A.; Cerutti, L.; Simanis, V.

submitted to the EMBL Data Library, April 1995

A:Description: The cdc15 gene is a key element in F-actin reorganisation at mitosis.

A:Reference number: S54119

A:Accession: S54119

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 'MLTKSLQ', 28-870, 'GKFIKEKL' <FA2>

A:Cross-references: EMBL:X86179

C:Genetics:

A:Gene: SPDB:SPAC20G8.05c; cdc15

A:Map position: 1

A:Introns: 27/3; 58/2; 871/1

C:Superfamily: SH3 homology

C:Keywords: mitosis; phosphoprotein

F:873-924/Domain: SH3 homology <SH3>

Query Match 43.6%; Score 44; DB 2; Length 927;

Best Local Similarity 47.1%; Pred. No. 69;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 LEEIAANSKDFPNMSET 18

Db 69 LQELAASSADIEFGST 85

RESULT 9

S67786

hypothetical protein YDL223c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D0843

C:Species: Saccharomyces cerevisiae

C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 29-Oct-1999

R;Accession: S67786

R;Rasmussen, S.W.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S67778

A;Accession: S67786

A;Molecule type: DNA

A;Residues: 1-1046 <RAS>

A;Cross-references: EMBL:Z74271; NID:q1431374; PID:e253363; PID:q1431375; GSPDB:GN000004;

A;Experimental source: strain S28C

C;Genetics:

A;Gene: MIPS:YDL223C

A;Map position: 4L

Query Match 43.6%; Score 44; DB 2; Length 1046;

Best Local Similarity 50.0%; Pred. No. 79;

Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 3 EETAAASKDFPNMSET 18

Db 733 QDIASDARDFTNNPET 748

RESULT 10

BABOH

peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine

N;Alternate names: aspartyl (asparaginyl) beta-hydroxylase

C;Species: Bos primigenius taurus (cattle)

C;Date: 31-Dec-1993 #sequence\_revision 10-Feb-1995 #text\_change 11-Jun-1999

R;Accession: A42969; A39470; C39470; S27948

R;Jia, S.; VanDusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern,

J. Biol. Chem. 267, 14322-14327, 1992

A;Title: cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.

A;Reference number: A42969; MUID:92332546

A;Accession: A42969

A;Molecule type: mRNA

A;Residues: 1-754 <JIA>

A;Cross-references: EMBL:M91213; NID:q162693; PIDN:AAA03563.1; PID:q162694

A;Experimental source: brain

A;Note: sequence extracted from NCBI backbone (NCBIP:108534)

R;Wang, Q.; VanDusen, W.J.; Petroski, C.J.; Garsky, V.M.; Stern, A.M.; Friedman, P.A.

J. Biol. Chem. 266, 14004-14010, 1991

A;Title: Bovine liver aspartyl beta-hydroxylase. Purification and characterization.

A;Reference number: A39470; MUID:91310689

A;Accession: A39470

A;Molecule type: protein

A;Residues: 289-328 <WAN>

A;Accession: B39470

A;Molecule type: protein

A;Residues: 615,'X',617-630,'XX',633-634,'X',636,'XX',639-641 <WA2>

A;Accession: C39470

A;Molecule type: protein

A;Residues: 311-347,'X',349,'X',351-373,'X',375-379,'X',381-382 <WA3>

C;Comment: This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the

C;Superfamily: peptidase-aspartate beta-dioxygenase; tetrahydropeptide repeat homology

C;Keywords: glycoprotein; oxidoreductase; intracellular #status predicted <INC>

F:2-56/Domain: intracellular #status predicted <INC>

F:57-78/Domain: transmembrane #status predicted <TRM>

F:289-754/Product: peptide-aspartate beta-dioxygenase, 56K form #status predicted <56K>

F:311-754/Product: peptide-aspartate beta-dioxygenase, 52K form #status predicted <52K>

F:337-370/Domain: tetrahydropeptide repeat homology <TT1>

F:371-404/Domain: tetrahydropeptide repeat homology <TT2>

F:13,96,466,702/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Best Local Similarity 43.1%; Score 43.5; DB 1; Length 754;

Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Qy 1 DLEETAAASKDFP-NMSE 17

Db 1 DLEETAAASKDFP-NMSE 17

RESULT 13

A64610

Db 82 DYEVLAKAKDFRYNLSE 99

RESULT 11

S65571

pattern formation protein GNOM - Arabidopsis thaliana

N;Alternate names: EMB30 protein

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 07-May-1999

C;Accession: S65571; S65572

R;Busch, M.; Mayer, U.; Juergens, G.

Mol. Gen. Genet. 250, 681-691, 1996

A;Title: Molecular analysis of the Arabidopsis pattern formation gene GNOM: gene stru

A;Reference number: S65571; MUID:96204508

A;Accession: S65571

A;Molecule type: DNA

A;Residues: 1-1451 <BUS>

A;Cross-references: EMBL:U36433; NID:gl209632; PID:gl209633

A;Accession: S65572

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 1-110,'I',112-866,'G',868-1451 <BUW>

A;Cross-references: EMBL:U36432; NID:gl209630; PID:gl209631

C;Genetics:

A;Gene: GNOM; EMB30

A;Introns: 246/3

Query Match

Best Local Similarity 43.1%; Score 43.5; DB 2; Length 1451;

Matches 12; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

Qy 1 DLEETAAANS-KDFPNMSET 18

Db 1302 DLEETAAAGSQDYRNMEGT 1320

RESULT 12

T29144

partial CBS - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000

C;Accession: T29144

R;Pauley, A.; Gattung, S.

submitted to the EMBL Data Library, July 1996

A;Description: The sequence of C. elegans cosmid K11C4.

A;Reference number: Z20577

A;Accession: T29144

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-5107 <PAU>

A;Cross-references: EMBL:U64854; PIDN:AA18318.1; GSPDB:GN000023; CESP:unc-68

A;Experimental source: strain Bristol N2; clone K11C4

C;Genetics:

A;Gene: CESP:unc-68

A;Map position: 5

A;Introns: 27/1; 64/3; 92/3; 127/1; 158/2; 1222/2; 1300/2; 1347/2; 1391/1; 1419/3; 15

/3; 3269/2; 3313/2; 3466/1; 3519/3; 3615/3; 3629/3; 3658/2; 3710/1; 3741/3; 3779/2; 3

C;Superfamily: ryanodine receptor; transcription initiation factor sigma region 1 hom

Query Match

Best Local Similarity 43.5%; Score 43.5; DB 2; Length 5107;

Matches 10; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

Qy 1 DLEETAAANS---KDFPNMSETDL 20

Db 3377 DLETVANNNTMYSQVNPVVDVL 3399

hypothetical protein hp0721 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 20-Jun-2000  
C:Accession: A64610  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalilak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467  
A:Accession: A64610  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-152 <TOM>  
A:Cross-references: GB:AE000585; GB:AE000511; NID:g2313845; PIDN:AAD07777.1; PID:g231385  
C:Superfamily: Helicobacter pylori hypothetical protein jhp0118

Query Match 42.6%; Score 43; DB 2; Length 152;  
Best Local Similarity 50.0%; Pred. No. 12;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 IAANSKDFPNMSETDL 20  
DB 16 VALNAKDFSKTSDLDL 31

## RESULT 14

A71904

hypothetical protein jhp0658 - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Jun-2000  
C:Accession: A71904  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.  
A:Reference number: A71800; MUID:99120557  
A:Accession: A71904  
A>Status: preliminary.  
A:Molecule type: DNA  
A:Residues: 1-152 <ARN>  
A:Cross-references: GB:AE001498; GB:AE001439; NID:g4155211; PIDN:AAD06236.1; PID:g415521  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0658  
C:Superfamily: Helicobacter pylori hypothetical protein jhp0118

Query Match 42.6%; Score 43; DB 2; Length 152;  
Best Local Similarity 50.0%; Pred. No. 12;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 IAANSKDFPNMSETDL 20  
DB 16 VALNAKDFSKTSDLDL 31

## RESULT 15

H70407

probable di-trans,poly-cis-decaprenylcistransferase (EC 2.5.1.31) aq\_1248 - Aquifex aeolicus  
N:Alternate names: decaprenyl diphosphate synthase  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 24-Sep-1999  
C:Accession: H70407  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: H70407



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:34:05 ; Search time 38.74 seconds  
(without alignments)  
17.685 Million cell updates/sec

Title: US-09-171-432A-45

Perfect score: 101

Sequence: 1 DLEEIAANSKDFPNMSETDL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	2226	1 POLG_HPAV4	P26581 hepatitis a
2	101	100.0	2226	1 POLG_HPAV8	P26582 hepatitis a
3	101	100.0	2227	1 POLG_HPAVH	P08617 hepatitis a
4	101	100.0	2227	1 POLG_HPAVL	P06441 hepatitis a
5	101	100.0	2227	1 POLG_HPAVM	P13901 hepatitis a
6	98	97.0	2230	1 POLG_HPAVS	P14553 simian hepa
7	97	96.0	2226	1 POLG_HPAV2	P26580 hepatitis a
8	46	45.5	194	1 Y041_THEAC	P57674 thermoplasm
9	44	43.6	740	1 GAG_SMRVH	P21411 squirrel mo
10	44	43.6	927	1 CC15_SCHPO	Q09822 schizosacch
11	43.5	43.1	754	1 ASPH_BOVIN	Q28056 bos taurus
12	43.5	43.1	1451	1 EM30_ARATH	Q42510 arabidopsis
13	43	42.6	231	1 UPRS_AQUAE	Q67291 aquifex aeo
14	41.5	41.1	1378	1 RPOB_CAMJE	Q46124 campylobact
15	41	40.6	346	1 RFAP_HAEIN	P45042 haemophilus
16	41	40.6	376	1 PT16_HUMAN	P35237 homo sapien
17	41	40.6	526	1 CLOS_CLOHI	P09870 clostridium
18	41	40.6	528	1 UD84_HUMAN	P06133 homo sapien
19	41	40.6	548	1 HLFB_VIBCH	P15492 vibrio chol
20	40	39.6	283	1 RUAP_SOYBN	P39657 glycine max
21	40	39.6	381	1 LDD_HAEIN	P46454 haemophilus
22	40	39.6	547	1 NLTP_HUMAN	P22307 homo sapien
23	40	39.6	824	1 NSFH_CAEEL	Q94392 caenorhabdi
24	39.5	39.1	244	1 P29_MYCPN	P75370 mycoplasma
25	39.5	39.1	521	1 VL2_MYCTU	Q07862 human papil
26	39.5	39.1	552	1 YMB1_HPV04	Q50684 mycobacteri
27	39	38.6	221	1 YSCL_YERPS	Q00928 versinia ps
28	39	38.6	223	1 YSCL_YEREN	Q01253 versinia en
29	39	38.6	253	1 PCRB_METJA	Q58647 methanococc
30	39	38.6	320	1 Y149_MYCPN	P75583 mycoplasma
31	39	38.6	396	1 LDD_ECOLI	P33232 escherichia
32	39	38.6	396	1 YJEB_YEAST	P47049 saccharomyc
33	39	38.6	438	1 GUX3_AGABI	P49075 agaricus bi

RESULT 1

ID	POLG_HPAV4	STANDARD:	PRT:	2226 AA.
AC	P26581;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)]			
DE	P3D (EC 2.7.7.48)]			
OS	Hepatitis A virus (strain 43c).			
OC	Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;			
OC	Hepatovirus.			
OX	NCBI_TaxID=12095;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91162758; PubMed=1705995;			
RA	Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,			
RA	Cromeans T., Jansen R.W.;			
RT	"Antigenic and genetic variation in cytopathic hepatitis A virus			
RT	variants arising during persistent infection: evidence for genetic			
RT	recombination."			
RL	J. Virol. 65:2056-2065(1991).			
CC	-!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,			
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,			
CC	VP3, AND VP4.			
CC	-!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.			
CC	-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL; M59809; AAA45469.1;			
DR	MEROPS; C03.005;			
DR	InterPro; IPR001205;			
DR	Pfam; PF00680; RNA_dep_RNA_pol; 1.			
DR	Pfam; PF00910; RNA_helicase; 1.			
KW	Polyprotein; Coat protein; Core protein; Core protein; Transferase;			
KW	RNA-directed RNA polymerase; Hydrolase; Thiol protease.			
FT	CHAIN 1 23			
FT	CHAIN 24 245			
FT	CHAIN 246 491			
FT	CHAIN 492 794			
FT	CHAIN 795 900			
FT	CHAIN 901 1087			
FT	CHAIN 1088 1422			
FT	CHAIN 1423 1495			
FT	CHAIN 1496 1518			
FT	CHAIN 1519 1737			
FT	CHAIN 1738 2226			



```

DR MEROPS; C03.005; -
DR InterPro: IPR000605; -
DR InterPro: IPR001205; -
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Hydrolase; Thiol protease;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 836
FT CHAIN 837 980
FT CHAIN 981 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1496
FT CHAIN 1497 1519
FT CHAIN 1520 1738
FT CHAIN 1739 2227
FT VARIANT 77 77
FT VARIANT 764 764
FT VARIANT 821 821
FT VARIANT 1052 1052
FT VARIANT 1062 1062
FT VARIANT 1118 1118
FT VARIANT 1151 1151
FT VARIANT 1163 1163
FT VARIANT 1277 1277
FT VARIANT 1500 1500
FT VARIANT 1805 1805
FT VARIANT 1930 1930
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AE8740A6 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 8.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 DLEEIAANSKDFPNMSETDL 950
|||||
RESULT 4
POLG_HPVLV STANDARD; PRT; 2227 AA.
AC P06441;
DT 01-JAN-1988 (rel. 06, Created)
DT 01-JAN-1988 (rel. 06, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain LA).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12099;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85190549; PubMed=2986127;
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,
RA Meriyweather J., van Nest G., Dina D.;
RT "Primary structure and gene organization of human hepatitis A virus.";
Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; K02990; AAA45472.1; -
DR MEROPS; C03.005; -
DR InterPro: IPR000605; -
DR InterPro: IPR001205; -
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Hydrolase; Thiol protease;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 836
FT CHAIN 837 980
FT CHAIN 981 1076
FT CHAIN 1077 1422
FT CHAIN 1423 1484
FT CHAIN 1485 1507
FT CHAIN 1508 1678
FT CHAIN 1679 2227
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 8.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 DLEEIAANSKDFPNMSETDL 950
|||||
RESULT 5
POLG_HPVLV STANDARD; PRT; 2227 AA.
AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;
AC Q81090; Q81091; Q81092; Q81093;
DT 01-JAN-1990 (rel. 13, Created)
DT 01-JAN-1990 (rel. 13, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain MBB).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88045071; PubMed=2823500;
RA Paul A.V., Tada H., der Helm K., Wessel T., Kiehn R., Wimmer E.,
RA Deinhardt F.;
RT "The entire nucleotide sequence of the genome of human hepatitis A
RT virus (isolate MBB).";
RL Virus Res. 8:153-171(1987).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20273: AAA45474.1; -.
CC PIR; JS0303; GNNYHB.
CC MEROPS; C03.005; -.
CC InterPro; IPR000605; -.
CC PIR; A30470; GNNYSA.
CC PIR; S04885; S04885.
CC MEROPS; C03.005; -.
CC InterPro; IPR001205; -.
CC PIR; PF00910; RNA_helicase; 1.
CC Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 836
FT CHAIN 837 980
FT CHAIN 981 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1496
FT CHAIN 1497 1519
FT CHAIN 1520 1738
FT CHAIN 1739 2227
SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;

Query Match 100.0%; Score 101; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 8, 1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
DB 931 DLEEIAANSKDFPNMSETDL 950

RESULT 6
POLG_HPAYS
ID POLG_HPAYS STANDARD; PRT; 2230 AA.
AC P14553;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Simian hepatitis A virus (strain AGM-27).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12102;
RN [1]
RP MEDLINE=91311420; PubMed=1649901;
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,
RA Purcell R.H.;
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome
RT structure and growth in cell culture with other HAV strains.";
RL J. Gen. Virol. 72:1677-1683(1991).
RN [2]
RP SEQUENCE OF 1750-2164 FROM N.A.
RX MEDLINE=89232168; PubMed=2541023;
RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;
RT "Variations in genome fragments coding for RNA polymerase in human
RT and simian hepatitis A viruses.";
RL FEBS Lett. 247:425-428(1989).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC EACH OF WHICH IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00924; BAA00766.1; -.
CC EMBL; X15461; CAA33490.1; -.
CC PIR; A30470; GNNYSA.
CC PIR; S04885; S04885.
CC MEROPS; C03.005; -.
CC InterPro; IPR001205; -.
CC PIR; PF00910; RNA_helicase; 1.
CC Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 27
FT CHAIN 28 249
FT CHAIN 250 495
FT CHAIN 496 795
FT CHAIN 796 984
FT CHAIN 985 1091
FT CHAIN 1092 1426
FT CHAIN 1427 1498
FT CHAIN 1499 1521
FT CHAIN 1522 1741
FT CHAIN 1742 2230
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;

Query Match 97.0%; Score 98; DB 1; Length 2230;
Best Local Similarity 95.0%; Pred. No. 2, 4e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
DB 935 DLEEIAANSKDFPNMSETDL 954

RESULT 7
POLG_HPAYS
ID POLG_HPAYS STANDARD; PRT; 2226 AA.
AC P26580;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 24a).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA Cromeans T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination.";
RL J. Virol. 65:2056-2065(1991).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M59810; AAA45468.1; -  
 DR MEROPS; C03.005; -  
 DR InterPro; IPR000605; -  
 DR InterPro; IPR001205; -  
 DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA\_helicase; 1.  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease;  
 FT CHAIN 1 23  
 FT CHAIN 24 245  
 FT CHAIN 246 491  
 FT CHAIN 492 794  
 FT CHAIN 795 900  
 FT CHAIN 901 1087  
 FT CHAIN 1088 1422  
 FT CHAIN 1423 1495  
 FT CHAIN 1496 1518  
 FT CHAIN 1519 1737  
 FT CHAIN 1738 2226  
 SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D684E2BF CRC64;

Query Match 96.0%; Score 97; DB 1; Length 2226;  
 Best Local Similarity 95.0%; Pred. No. 3.4e-07;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEETAANSKDPNNSETDL 20  
 DB 931 DLEETAANSKDPNNSETDL 950

RESULT 8  
 Y041\_THEAC  
 ID Y041\_THEAC STANDARD; PRT; 194 AA.  
 AC P57674;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PROTEIN TA0041.  
 GN TA0041.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;  
 OC Thermoplasma.  
 OX NCBI\_TaxID=2303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=20479972; PubMed=11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
 acidophilum";  
 RL Nature 407:508-513(2000).  
 CC -!- SIMILARITY: BELONGS TO THE UPF0129 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AL445063; CAC11190.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 194 AA; 21854 MW; 6F7CC0D844F9FF4F CRC64;

Query Match 45.5%; Score 46; DB 1; Length 194;  
 Best Local Similarity 50.0%; Pred. No. 2.8;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 LEEETAANSKDPNNSETD 19  
 DB 92 VEETAARKTGDLNLSQTD 109

RESULT 9  
 GAG\_SMRVH  
 ID GAG\_SMRVH STANDARD; PRT; 740 AA.  
 AC P21411;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 01-MAY-1991 (Rel. 18, Last annotation update)  
 DE GAG POLYPROTEIN [CONTAINS: CORE PROTEINS P19, P16; PROBABLE CORE  
 DE PROTEIN P35, P10].  
 GN GAG.  
 OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).  
 OC Viruses; Retroviral viruses; Retroviridae; Betaretrovirus.  
 OX NCBI\_TaxID=111856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89073750; PubMed=3201749;  
 RA Oda T., Ikeda S., Watanabe S., Hatsushika M., Akiyama K.,  
 RA Mitsunobu F.;  
 RT "Molecular cloning, complete nucleotide sequence, and gene structure  
 RT of the provirus genome of a retrovirus produced in a human  
 RT lymphoblastoid cell line";  
 RL Virology 167:468-476(1988).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 43.6%; Score 44; DB 1; Length 740;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 10; Conservative 4; Mismatches 4; Indels 2; Gaps 1;  
 QY 1 DLEETAA--NSKDPNNSET 18  
 DB 214 DLEETAAQYNNPDWPLNT 233

RESULT 10  
 CC15\_SCHPO  
 ID CC15\_SCHPO STANDARD; PRT; 927 AA.  
 AC C09822; O14365;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE CELL DIVISION CONTROL PROTEIN 15.  
 GN CDC15 OR SPAC20G8.05C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OX NCBI\_TaxID=4896;



```
QY 1 DLEEIAANS-KDFPNMSET 18
    || |||| | ||: || |
Db 1302 DLLEIAAGSQKDYRNMECT 1320
```

```
[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN-NCTC 11168;
RX  MEDLINE=20150912; PubMed=10698204;
RA  Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA  Basham D., Chillingworth T., Davies R.M., Feltham T., Holt R.D.,
RA  Jorgensen F., Karlyshev A.V., Moule S., Pallen M.J., Parkhill J.,
RA  Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA  Whitehead S., Barrell B.G.;
RT  "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT  reveals hypervariable sequences."
RL  Nature 403:665-668(2000).
[2]
RN  SEQUENCE OF 338-1031 FROM N.A.
RP  MEDLINE=9608494; PubMed=7489896;
RA  Bustamante V.H., Puente J.L., Sanchez-Lopez F., Bobadilla M.,
RA  Calva E.;
RT  "Identification of Campylobacter jejuni and C. coli using the rpoB
RT  gene and a cryptic DNA fragment from C. jejuni."
RL  Gene 165:1-8(1995).
CC  -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC  OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC  SUBSTRATES.
CC  -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE +
CC  RNA(N) -> RNA(N) + PPi
CC  -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC  ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC  BETA' CHAIN.
CC  -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AL139075; CAB75116.1; -
DR  EMBL; X77304; CAAS4509.1; -
DR  InterPro; IPR001572; -
DR  Pfam; PF00562; RNA_POL_B; 1.
DR  PROSITE; PS01166; RNA_POL_BETA; 1.
KW  Transferase; Transcription; DNA-directed RNA polymerase.
FT  CONFLICT 338 347 NDLANGVDA -> MFWLMALMQP (IN REF. 2).
FT  CONFLICT 558 558 A -> R (IN REF. 2).
FT  CONFLICT 671 671 C -> S (IN REF. 2).
FT  CONFLICT 691 691 A -> R (IN REF. 2).
SQ  SEQUENCE 1378 AA; 155915 MW; AB7467C305028EB5 CRC64;

Query Match 41.18; Score 41.5; DB 1; Length 1378;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

QY 2 LEEAANSKDFPNMSETDL 20
   :||| :||| :||| :|||
DB 858 IEEI---TKDIPNKEEDV 873

RESULT 15
ID RFAF_HAEIN STANDARD; PRT; 346 AA.
AC P45042;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ADP-HEPTOSE--LPS HEPTOSYLTRANSFERASE II.
GN RFAF OR H1105.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
```

```
[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN-RD / KW20 / ATCC 51907;
RX  MEDLINE=95350630; PubMed=7542800;
RA  Fieischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA  Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA  Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA  Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA  Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA  Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA  Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA  Venter J.C.;
RT  "Whole-genome random sequencing and assembly of Haemophilus
RT  influenzae Rd."
RL  Science 269:496-512(1995).
CC  -!- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U32790; AAC22760.1; -
DR  TIGR; H11105; -
DR  InterPro; IPR002201; -
DR  Pfam; PF01075; Heptosyltransf; 1.
KW  Lipopolysaccharide biosynthesis; Transferase.
SQ  SEQUENCE 346 AA; 38847 MW; 14D0E11C6773FCA9 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 346;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 LEEAANSKDFPNM 15
   |::|::|::|
DB 116 LNDLRANKDYPMM 129

Search completed: April 24, 2001, 16:34:08
Job time: 564 sec
```

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:33:17 ; Search time 113.97 Seconds  
(without alignments)  
20.568 Million cell updates/sec

Title: US-09-171-432A-45  
Perfect score: 101  
Sequence: 1 DLEEIAANSKDFPNMSETDL 20

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	101	100.0	1124	14 Q84780	Q84780 hepatitis a
2	101	100.0	1161	14 Q05794	Q05794 hepatitis a
3	101	100.0	2216	14 Q9WMA2	Q9WMA2 hepatitis a
4	101	100.0	2218	14 Q67824	Q67824 hepatitis a
5	101	100.0	2227	14 Q67825	Q67825 hepatitis a
6	101	100.0	2227	14 Q67826	Q67826 hepatitis a
7	101	100.0	2227	14 Q9WMA4	Q9WMA4 hepatitis a
8	101	100.0	2227	14 Q9WMA3	Q9WMA3 hepatitis a
9	101	100.0	2227	14 Q9WMA1	Q9WMA1 hepatitis a
10	101	100.0	2227	14 Q9WMA0	Q9WMA0 hepatitis a
11	101	100.0	2227	14 Q91PH5	Q91PH5 hepatitis a
12	95	94.1	2218	14 Q67817	Q67817 hepatitis a
13	94	93.1	184	14 Q87032	Q87032 simian hepa
14	91	90.1	2227	14 Q9WM99	Q9WM99 hepatitis a
15	49	48.5	400	5 O16961	O16961 caenorhabdi
16	46	45.5	424	5 Q9VW80	Q9VW80 drosophila
17	45.5	45.0	930	10 Q92VD5	Q92VD5 arabidopsis
18	45.5	45.0	1621	5 O01900	O01900 caenorhabdi
19	44.5	44.1	1673	5 Q9V599	Q9V599 drosophila

20	44	43.6	1046	3 Q07653	Q07653 saccharomyc
21	43.5	43.1	225	4 Q9NR10	Q9NR10 homo sapien
22	43.5	43.1	5071	5 P91905	P91905 caenorhabdi
23	43.5	43.1	5107	5 Q94279	Q94279 caenorhabdi
24	43	42.6	152	2 Q25423	Q25423 helicobacte
25	43	42.6	152	2 Q9ZLC2	Q9ZLC2 helicobacte
26	43	42.6	210	14 Q9JDD9	Q9JDD9 human immun
27	43	42.6	210	14 Q9JDC4	Q9JDC4 human immun
28	43	42.6	294	9 O48432	O48432 lactobacill
29	43	42.6	350	10 Q9LEL5	Q9LEL5 coptis japo
30	43	42.6	454	10 O80668	O80668 arabidopsis
31	43	42.6	598	2 O83350	O83350 treponema p
32	43	42.6	626	2 Q9S3K2	Q9S3K2 eubacterium
33	43	42.6	627	8 Q9TLL1	Q9TLL1 nephroseimi
34	43	42.6	1072	2 Q9K6N1	Q9K6N1 bacillus ha
35	43	42.6	1201	5 O16844	O16844 drosophila
36	43	42.6	1201	5 Q9V4K0	Q9V4K0 drosophila
37	42.5	42.1	913	5 O01922	O01922 caenorhabdi
38	42.5	42.1	4717	3 Q94248	Q94248 schizosacch
39	42	41.6	68	14 O11619	O11619 human immun
40	42	41.6	164	14 Q87010	Q87010 subterranea
41	42	41.6	289	5 Q9XU05	Q9XU05 caenorhabdi
42	42	41.6	365	2 P71278	P71278 enterobacte
43	42	41.6	365	2 Q9JY60	Q9JY60 neisseria m
44	42	41.6	365	2 Q9JY44	Q9JY44 neisseria m
45	42	41.6	402	5 Q9W1D2	Q9W1D2 drosophila

#### ALIGNMENTS

RESULT 1

Q84780 PRELIMINARY: PRT: 1124 AA.

AC Q84780:

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS)

DE (FRAGMENT).

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

[1]

RP SEQUENCE FROM N.A.

RA Ovchinnikov Y.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,

RA Rokhlina T.O., Chizhikov V.E., Petrov N.A., Prikhod'ko G.G.,

RA Blinov V.M., Vasilenko S.K., Sandakhchiev L.S., Kusov Y.Y.,

RL Grabko V.I., Fleer G.P., Balyan M.S., Drozdov S.G.;

RL Dokl. Biochem. 285:379-383(1986).

DR EMBL; X04200; CAA27797.1; -

DR EMBL; A11312; CAA00953.1; -

KW Nonstructural protein.

FT NON\_TER 1124 1124

SQ SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

Query Match 100.0%; Score 101; DB 14; Length 1124;  
Best local Similarity 100.0%; Pred. No. 1.le-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20  
|||||

Db 888 DLEEIAANSKDFPNMSETDL 907

RESULT 2

Q05794 PRELIMINARY: PRT: 1161 AA.

ID Q05794

AC Q05794; Q67801; Q67802; Q67803; Q67804; Q67805; Q67806;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE GENOME POLYPEPTIDE (COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
 P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
 P3D (EC 2.7.7.48)) (FRAGMENT).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilenko S.K.,  
 RA Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,  
 RA Balayan M.S.;  
 RL Mol. Gen. Microbiol. Virol. 6:129-133(1987).  
 CC -1- SUBUNT: THE VIRUS CA5D1S IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
 CC VP3, AND VP4.  
 DR EMBL; X15464; CAA33492.1; -;  
 DR INTERPRO: IPR000408; -;  
 DR PROSITE: PS00626; RCC1\_2; UNKNOWN\_1.  
 KW Polypeptide; Coat protein; Core protein; RNA-directed RNA polymerase;  
 KW Hydrolase; Thiol protease.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
 FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
 FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
 FT CHAIN 795 900 CORE PROTEIN P2A.  
 FT CHAIN 901 1087 CORE PROTEIN P2B.  
 FT CHAIN 1088 >1161 CORE PROTEIN P2C.  
 FT NON\_TER 1161 1161  
 SQ SEQUENCE 1161 AA; 131131 MW; 38BE93789PEC3400 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 1161;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLEEIAANSKDFPNMSETDL 20  
 |||||  
 DB 925 DLEEIAANSKDFPNMSETDL 944

RESULT 3  
 Q67825 PRELIMINARY; PRT; 2216 AA.  
 ID Q67825  
 AC Q67825  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE POLYPEPTIDE.  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AH3;  
 RA Fujiwara K.;  
 RL "Hepatitis A virus."  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB020566; BAA35104.1; -;  
 DR INTERPRO: IPR000408; -;  
 DR INTERPRO: IPR000605; -;  
 DR INTERPRO: IPR001205; -;  
 DR INTERPRO: IPR001643; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol.1.  
 DR PRINTS; PR00910; RNA\_helicase.1.  
 DR PROSITE; PR00918; CALICIVIRUSNS.  
 KW Polypeptide.  
 SQ SEQUENCE 2216 AA; 250209 MW; 1A9D93FEC21FBE82 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2216;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLEEIAANSKDFPNMSETDL 20  
 |||||  
 DB 931 DLEEIAANSKDFPNMSETDL 950

RESULT 4  
 Q67824 PRELIMINARY; PRT; 2218 AA.  
 ID Q67824  
 AC Q67824  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE GBM/FRHK RNA.  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GBM;  
 RX MEDLINE=94076453; PubMed=8254770;  
 RA Graff J., Normann A., Feinstein S.M., Flehmig B.;  
 RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
 RT with two cell culture-adapted variants.";  
 RL J. Virol. 68:548-554(1994).  
 DR EMBL; X75214; CAA53024.1; -;  
 DR INTERPRO: IPR000408; -;  
 DR INTERPRO: IPR000605; -;  
 DR INTERPRO: IPR001205; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol.1.  
 DR PFAM; PF00910; RNA\_helicase.1.  
 DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.  
 FT CHAIN 1 785 P1 STRUCTURAL PROTEINS.  
 FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEINS.  
 FT CHAIN 1417 2218 P3 NONSTRUCTURAL PROTEINS.  
 SQ SEQUENCE 2218 AA; 250502 MW; CA72DF0922104C0E CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2218;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLEEIAANSKDFPNMSETDL 20  
 |||||  
 DB 925 DLEEIAANSKDFPNMSETDL 944  
 RESULT 5  
 Q67825 PRELIMINARY; PRT; 2227 AA.  
 ID Q67825  
 AC Q67825  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE GBM/WT RNA.  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GBM;  
 RX MEDLINE=94076453; PubMed=8254770;  
 RA Graff J., Normann A., Feinstein S.M., Flehmig B.;  
 RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
 RT with two cell culture-adapted variants.";  
 RL J. Virol. 68:548-554(1994).  
 DR EMBL; X75215; CAA53025.1; -;



```
DR INTERPRO; IPR000408; -
DR INTERPRO; IPR000605; -
DR INTERPRO; IPR001205; -
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PROSITE; PS00626; RC1_2; UNKNOWN_1.
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.
FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 2227 AA; 251563 MW; 4C4D79D352F936B4 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 DLEEIAANSKDFPNMSETDL 950

RESULT 6
Q67826 PRELIMINARY; PRT; 2227 AA.
AC Q67826;
RC STRAIN-GBM;
RX MEDLINE-94076453; PubMed-8254770;
RA Graff J., Normann A., Feinstein S.M., Flehmig B.;
RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison
  with two cell culture-adapted variants.";
RL J. Virol. 68:548-554(1994).
DR EMBL; X75216; CAA53026.1; -
DR INTERPRO; IPR000408; -
DR INTERPRO; IPR001205; -
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PROSITE; PS00626; RC1_2; UNKNOWN_1.
FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.
FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.
FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 2227 AA; 251496 MW; 488CB7C962319457 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 DLEEIAANSKDFPNMSETDL 950

RESULT 7
Q9WMA4 PRELIMINARY; PRT; 2227 AA.
AC Q9WMA4;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID-12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH1;
RA Fujiwara K.;
RT "Hepatitis A virus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AB020564; BAA35102.1; -
DR INTERPRO; IPR000408; -
DR INTERPRO; IPR001205; -
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSN.
DR PROSITE; PS00626; RC1_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251304 MW; 0DEF6D2AEC29C0CE CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 DLEEIAANSKDFPNMSETDL 950

RESULT 8
Q9WMA3 PRELIMINARY; PRT; 2227 AA.
AC Q9WMA3;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID-12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AH2;
RA Fujiwara K.;
RT "Hepatitis A virus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AB020565; BAA35103.1; -
DR INTERPRO; IPR000408; -
DR INTERPRO; IPR001205; -
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSN.
DR PROSITE; PS00626; RC1_2; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 2227 AA; 251440 MW; E04F846CEC7F50FD CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
Db 931 DLEEIAANSKDFPNMSETDL 950

RESULT 9
Q9WMA1 PRELIMINARY; PRT; 2227 AA.
AC Q9WMA1;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
```

```

Db      931 DLEEIAANSKDFPNMSETDL 950
|||||
RESULT  11
Q91FH5 PRELIMINARY; PRT; 2227 AA.
ID Q91FH5;
AC Q91FH5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYPROTEIN PRECURSOR.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HAF-203;
RA Baptista M.L., Silva M., de Lima M.A., Yoshida C.F., Gaspar A.M.,
RA Pires Lopes M.O., Galler R.;
RT "Nucleotide sequence of the HAF-203 hepatitis A virus strain isolated
RT in Brazil and expression of the VP1 gene in a bacterial system.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF268396; AAF80114.1; -.
SQ SEQUENCE 2227 AA; 251432 MW; 81913AEC8A04200 CRC64;

Query Match 100.0%; Score 101; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEEIAANSKDFPNMSETDL 20
|||||
Db      931 DLEEIAANSKDFPNMSETDL 950

RESULT  12
Q67817 PRELIMINARY; PRT; 2218 AA.
ID Q67817;
AC Q67817;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE COMPLETE GENOME.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F.G.;
RA MEDLINE=95381623; PubMed=7653108;
RX Beneduce F., Pisani G., Divizia M., Pana A., Morace G.;
RT "Complete nucleotide sequence of a cytopathic hepatitis A virus strain
RT isolated in Italy.";
RL Virus Res. 36:299-309(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F.G.;
RA Morace G.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83302; CAA58281.1; -.
DR INTERPRO; IPR000408; -.
DR INTERPRO; IPR000605; -.
DR INTERPRO; IPR001205; -.
DR PFAM; PF00680; RNA_dep_RNA_pol.1.
DR PFAM; PF00910; RNA_helicase.1.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 2218 AA; 250476 MW; 813B21D3E4E533CA CRC64;

Query Match 94.1%; Score 95; DB 14; Length 2218;

```

Best Local Similarity 100.0%; Pred. No. 2.1e-06;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LEEIAANSKDFPNMSETDL 20  
DB 926 LEEIAANSKDFPNMSETDL 944

RESULT 13

ID Q87092 PRELIMINARY; PRT; 184 AA.  
AC Q87092;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE SEGMENT 2A-ENCODED PROTEIN (FRAGMENT).  
OS Simian hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90012354; PubMed=2552172;  
RA Brown E.A.; Jansen R.W.; Lemon S.M.;  
RT "Characterization of a simian hepatitis A virus (HAV): antigenic and  
RT genetic comparison with human HAV."  
RL J. Virol. 63:4932-4937(1989).  
DR EMBL; M34085; AAA47483.1; -  
FT NON\_TER 1 184  
FT NON\_TER 184 184  
SQ SEQUENCE 184 AA; 21322 MW; 9FA851FA964A88A8 CRC64;

Query Match 93.1%; Score 94; DB 14; Length 184;  
Best Local Similarity 90.0%; Pred. No. 2.1e-07;  
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEEIAANSKDFPNMSETDL 20  
DB 140 LEEIAANSKDFPNMSETDL 159

RESULT 14

ID Q9WM99 PRELIMINARY; PRT; 2227 AA.  
AC Q9WM99;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FH3;  
RA Fujiwara K.;  
RT "Hepatitis A virus."  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020569; BAA35107.1; -  
DR INTERPRO; IPR000408; -  
DR INTERPRO; IPR000605; -  
DR INTERPRO; IPR001205; -  
DR INTERPRO; IPR001643; -  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSN.  
DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2227 AA; 251387 MW; 817640187672A23D CRC64;

Query Match 90.1%; Score 91; DB 14; Length 2227;  
Best Local Similarity 90.0%; Pred. No. 9.3e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LEEIAANSKDFPNMSETDL 20  
DB 931 LEEIAANSKDFPNMSETDL 950

RESULT 15

ID O16961 PRELIMINARY; PRT; 400 AA.  
AC O16961;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE R11G11.1 PROTEIN.  
GN R11G11.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R.; Alnscough R.; Anderson K.; Baynes C.; Berks M.;  
RA Bonfield J.; Burton J.; Connell M.; Copsey T.; Cooper J.; Coulson A.;  
RA Craxton M.; Dear S.; Du Z.; Durbin R.; Favello A.; Fulton L.;  
RA Gardner A.; Green P.; Hawkins T.; Hillier L.; Jier M.; Johnston L.;  
RA Jones M.; Kershaw J.; Kirsten J.; Laister N.; Latreille P.;  
RA Lightning J.; Lloyd C.; McMurray A.; Mortimore B.; O'Callaghan M.;  
RA Parsons J.; Percy C.; Rifken L.; Roopra A.; Saunders D.; Showkeen R.;  
RA Smaldon N.; Smith A.; Sonhammer E.; Staden R.; Sulston J.;  
RA Thierry-Mieg J.; Thomas K.; Vaubin M.; Vaughan K.; Waterston R.;  
RA Watson A.; Weinstock L.; Wilkinson-Sproat J.; Wohldman P.; III of C.  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans."  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Davidson S.; Wohldmann P.; Bauer C.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF022976; AAC69076.1; -  
DR INTERPRO; IPR000536; -  
DR PFAM; PF00104; hormone\_rec; 1.  
DR PFAM; PF00105; zf-C4; 2.  
DR PRINTS; PR00047; STROIDFINGER.  
SQ SEQUENCE 400 AA; 46050 MW; CD32B223C12B2068 CRC64;

Query Match 48.5%; Score 49; DB 5; Length 400;  
Best Local Similarity 47.4%; Pred. No. 7.3;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LEEIAANSKDFPNMSETDL 20  
DB 164 LENTSTNMKDFPHSKEDV 182

Search completed: April 24, 2001, 16:33:20  
Job time: 581 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:23:33 ; Search time 121.9 Seconds  
(without alignments)  
9.379 Million cell updates/sec

Title: US-09-171-432A-46

Perfect score: 96

Sequence: 1 KINLADRLGLSGVQEIKEQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0401.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	20	W42929	Immunogenic Hepati
2	96	100.0	25	W42969	Immunogenic Hepati
3	96	100.0	2227	P60066	Sequence of viral
4	96	100.0	2227	11 R05697	Attenuated hepatit
5	96	100.0	2227	18 W34074	Hepatitis A virus
6	96	100.0	2227	21 B18607	Amino acid sequenc
7	96	100.0	2227	21 B18608	Amino acid sequenc
8	96	100.0	2227	21 B18609	Amino acid sequenc
9	41	42.7	332	20 Y37809	Protein which is s
10	41	42.7	366	19 W60855	Mouse CD14 protein
11	41	42.7	366	21 Y53879	A murine CD14 prot

12	41	42.7	708	19 W69844	Amino acid sequenc
13	41	42.7	748	19 W50143	Cyanobacterial phy
14	40	41.7	332	11 R03167	Amino acid sequenc
15	40	41.7	423	21 B01670	Yeast type II topo
16	40	41.7	440	20 Y34626	Chlamydia pneumoni
17	40	41.7	643	20 Y16797	Amino acid sequenc
18	40	41.7	1186	20 Y16796	Amino acid sequenc
19	40	41.7	1227	19 W44321	Bacillus thuringie
20	40	41.7	1227	20 Y31990	Chimeric CryIIB ins
21	40	41.7	1227	22 B19950	Bacillus thuringie
22	40	41.7	1228	15 R50955	Bacillus thuringie
23	40	41.7	1229	15 R54074	CryET5. Bacillus
24	40	41.7	1229	18 W35259	Bacillus thuringie
25	40	41.7	1229	18 W17699	CryET5. Bacillus
26	40	41.7	1229	20 Y30923	B. thuringiensis c
27	40	41.7	1229	20 W87633	CryET5 protein seq
28	40	41.7	1372	19 W56473	Protein with Rho p
29	39	40.6	87	21 C59693	Zea mays protein f
30	39	40.6	378	19 W98395	H. pylori GPO 117
31	39	40.6	388	20 Y31897	Corn cyclin delta-
32	39	40.6	698	20 Y36836	Protein involved i
33	39	40.6	748	21 B26592	Synechocystis sp p
34	38.5	40.1	745	22 B65611	Novel protein kina
35	38	39.6	44	16 R80835	Human CD40 ligand
36	38	39.6	111	21 G04065	Human secreted pro
37	38	39.6	185	18 W25771	Mouse mb52 Cl. Mu
38	38	39.6	218	22 B63628	Human gastric canc
39	38	39.6	226	15 R49548	Predicted sequence
40	38	39.6	261	14 R36701	CD40-L. Homo sapi
41	38	39.6	261	15 R53969	Human CD40-L type
42	38	39.6	261	15 R57469	CD40 ligand. Synt
43	38	39.6	261	16 R85486	Human CD40 ligand.
44	38	39.6	261	16 R76125	Human CD40-L. Hom
45	38	39.6	261	18 W34669	Amino acid sequenc

#### ALIGNMENTS

RESULT 1

W42929

ID W42929 standard; peptide; 20 AA.

XX

AC W42929;

XX

DT 28-APR-1998 (first entry)

XX

DE Immunogenic Hepatitis A virus peptide YK-1331.

XX

KW Immunogenic peptide; immunogenic epitope; P2A protein;

KW immune response; antibody.

XX

OS Synthetic.

OS Hepatitis A virus.

XX

PN WO9740147-A1.

XX

PD 30-OCT-1997.

XX

PF 18-APR-1997; 97WO-US06891.

XX

PR 19-APR-1996; 96US-0015644.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI Fields HA, Khudyakov YE;

DR WPI; 1997-535831/49.

XX

PT Immunogenic Hepatitis A virus (HAV) peptide(s) - used to induce an

PT immune response to HAV in a mammal or to detect the presence of

PT antibodies against HAV in a mammal

XX



XX	Sequence	2227 AA;	100.0%; Score 96; DB 7; Length 2227;	0; Gaps	0;
XX	Query Match		100.0%; Score 96; DB 7; Length 2227;		
XX	Best Local Similarity	100.0%; Pred. No. 7.8e-08;			
XX	Matches	20; Conservative	0; Mismatches	0; Indels	0; Gaps
QY	1 KINLADRMIGLSGVQEIKEQ 20				
DB	961 kinladrmiglsgvqeiqeq 980				
RESULT	4				
ID	R05697	standard; protein; 2227 AA.			
XX	AC	R05697;			
XX	DT	15-AUG-1990 (first entry)			
XX	DE	Attenuated hepatitis A virus.			
XX	DE	Hepatitis A virus; vaccine; attenuated.			
XX	DE	Hepatitis A virus, strain HM-175.			
XX	DE	Key	Location/Qualifiers		
XX	DE	1..23	/label=VP4 = 1A		
XX	DE	24..245	/label=VP2 = 1B		
XX	DE	246..491	/label=VP3 = 1C		
XX	DE	492..791	/label=VP1 = 1D		
XX	DE	792..980	/label=2A		
XX	DE	981..1087	/label=2B		
XX	DE	1088..1422	/label=2C		
XX	DE	1423..1496	/label=3A		
XX	DE	1497..1519	/label=3B = VPg		
XX	DE	1520..1738	/label=3C		
XX	DE	1739..2227	/label=3D		
XX	US4894228-A.				
XX	16-JAN-1990.				
XX	12-JUL-1988;	88US-0217824.			
XX	12-JUL-1988;	88US-0217824.			
XX	12-JUL-1988;	88US-0652967.			
XX	(USSH ) US DEPT HEALTH & HUMAN.				
XX	Purcell RH, Ticehurst JR, Cohen I, Emerson SU, Feinstein SM;				
XX	Daemer RJ, Gust ID;				
XX	WPI; 1990-075557/10.				
XX	N-PSDB; Q03512.				
XX	Vaccine against hepatitis A virus infection - comprises novel				
XX	attenuated hepatitis A virus strain.				
XX	Claim 1; Fig 1; 18pp; English.				

CC	(HAV) strain HM-174. The sequence is modified to produce HAV which are adapted to growth in the human fibroblast-like cell line MRC-5.
CC	The HAV is able to propagate in MRC-5 cells and retain appropriate attenuation. It is useful as a live vaccine for prophylaxis of hepatitis A in humans and other primates.
XX	
SQ	Sequence 2227 AA;
 Query Match 100.0%; Score 96; DB 21: Length 2227; Best Local Similarity 100.0%; Pred. No. 7.8e-08; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 KINLADRMGLSGVQEIKEQ 20 
Dd	961 kinladrmglsgvqeikeq 980 
 RESULT 7 BI8608 ID BI8608 standard; Protein; 2227 AA. XX AC BI8608; XX XX DT 15-JAN-2001 (first entry) XX XX DE Amino acid sequence of passage 35 Hepatitis A virus called P-35. KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection; KW P-35 virus. XX OS Hepatitis A virus. OS OS PN US6113912-A. XX XX XX 05-SEP-2000. PD PD XX 07-JUN-1995; 95US-0475886. XX XX PR 18-SEP-1992; 92US-0947338. PR 17-SEP-1993; 93WO-US08610. PR 10-MAR-1995; 95US-0397232. XX XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES. XX XX PI D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW; XX XX WPI; 2000-586464/55. DR N-PSDB; A75477. XX XX PPT Novel live hepatitis A virus adapted to growth in human fibroblast cell line useful as vaccine for protecting humans against hepatitis A virus infection, has modified genome compared to wild type - PPT Disclosure; Columns 67-78; 72pp; English. XX XX PS The present sequence is derived from passage 35 of a wild type hepatitis A virus (HAV) strain HM-174. The resulting virus is designated P-35 virus. The sequence is modified to produce HAV which are adapted to growth in the human fibroblast-like cell line MRC-5. CC The HAV is able to propagate in MRC-5 cells and retain appropriate attenuation. It is useful as a live vaccine for prophylaxis of hepatitis A in humans and other primates. XX XX SQ Sequence 2227 AA;	
 Query Match 100.0%; Score 96; DB 21: Length 2227; Best Local Similarity 100.0%; Pred. No. 7.8e-08; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 KINLADRMGLSGVQEIKEQ 20 



Db 961 kinladrmlgslsgvqeikey 980

RESULT 8

BL8609

ID BL8609 standard; Protein: 2227 AA.

AC BL8609;

XX

DT 15-JAN-2001 (first entry)

XX

DE Amino acid sequence of live attenuated Hepatitis A virus 4380.

XX

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

KW HAV 4380.

XX

OS Hepatitis A virus.

XX

PN US6113912-A.

XX

PD 05-SEP-2000.

XX

XX 07-JUN-1995; 95US-0475886.

PF

XX 18-SEP-1992; 92US-0947338.

PR

PR 17-SEP-1993; 93WO-US08610.

PR

XX 10-MAR-1995; 95US-0397232.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX

XX D'Hondt E, Purcell RH, Emerson SU, Funkhouser AW;

PI

XX WPI; 2000-586464/55.

DR

DR N-PSDB; A75478.

XX

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus

PT infection, has modified genome compared to wild type

XX

PS Disclosure; Columns 93-104; 72pp; English.

XX

CC The present sequence is derived from a live attenuated hepatitis A

CC virus (HAV) of the invention, designated HAV 4380. The sequence is

CC produced by modifying wild type HAV strain HM-174. The HAV of the

CC invention are adapted to growth in the human fibroblast-like cell

CC line MRC-5. The HAV is able to propagate in MRC-5 cells and retain

CC appropriate attenuation. It is useful as a live vaccine for prophylaxis

CC of hepatitis A in humans and other primates.

XX

SQ Sequence 2227 AA;

Query Match 100.0%; Score 96; DB 21; Length 2227;

Best Local Similarity 100.0%; Pred. No. 7.8e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGVQEIKEY 20

|||||

Db 961 kinladrmlgslsgvqeikey 980

RESULT 9

Y37809

ID Y37809 standard; Protein: 332 AA.

AC Y37809;

XX

XX 07-OCT-1999 (first entry)

DT

DE Protein which is specific to Chlamydia trachomatis.

XX

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;

KW

nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

Chlamydia trachomatis.

WO9928475-A2.

PN

XX 10-JUN-1999.

PD

XX 27-NOV-1998; 98WO-IB01939.

PF

XX 04-NOV-1998; 98US-0107077.

PR

PR 28-NOV-1997; 97FR-0015041.

PR

XX 17-DEC-1997; 97FR-0016034.

XX

PA (GEST ) GENSET.

XX

XX Griffais R;

PI

XX WPI; 1999-371125/31.

DR

XX Genome sequence of Chlamydia trachomatis

XX

PS Disclosure; Page 1384-1385; 1755pp; English.

XX

CC Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of

CC Chlamydia trachomatis (see 201425). The polypeptides can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia

CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,

CC perinephritis, bartholinitis; pneumonia; breast feeding infants;

CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.

XX

SQ Sequence 332 AA;

Query Match 42.7%; Score 41; DB 20; Length 332;

Best Local Similarity 44.4%; Pred. No. 33;

Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 NLADRLGLSGVQEIKEY 20

|||||

Db 259 needrfialsgaylsee 276

RESULT 10

W60855

ID W60855 standard; Protein: 366 AA.

XX

AC W60855;

XX

DT 01-OCT-1998 (first entry)

DT

DE Mouse CD14 protein.

XX

XX CD14; B cell activator; bovine lactation-associated immunotropic protein;

KW LAIT; T cell immunodeficiency; X-linked hyper-IgM syndrome; allergy;

KW common variable immunodeficiency; X-linked agammaglobulinaemia; vaccine;

XX infant feeding formulae; mouse.

XX

OS Mus sp.

XX

XX WO9822580-A2.

PN

XX 28-MAY-1998.

PD

XX 18-NOV-1997; 97WO-CA00880.

PF

XX 18-NOV-1996; 96US-0746883.

PR



DR WPI: 1998-377650/32.  
 XX N-PSDB; V42649.

PT New isolated vesicle secretion associated poly:peptide(s) - used to  
 PT develop products for treating e.g. effective disorder.  
 PT neurodegenerative disease, hormone imbalances, immune system  
 PT disorders or tumours

XX Claim 1a; Pages 85-87; 133pp; English.

PS The present sequence represents a p71 protein. This protein is present  
 CC the secretion associated 17S (SA-17S) complex. Eight proteins form the  
 CC SA-17S complex, which binds a syntaxin-containing (SC) complex. The  
 CC SA-17S polypeptides and nucleotide sequences encoding them can be used  
 CC for screening for compounds which modulate vesicular release involved in  
 CC synaptic transmission and other secretory processes. Compounds which  
 CC enhance binding between the SA-17S and SC complexes may be used to treat  
 CC an affective disorder such as depression, manic-depressive disorders and  
 CC anxiety disorders, or a neurodegenerative disease such as Parkinson's  
 CC disease or Huntington's disease. Compounds which inhibit binding between  
 CC the SA-17S and SC complexes may be used to treat a disorder of thought,  
 CC such as schizophrenia, or for anaesthesia. The compounds can also be used  
 CC to intervene in the endocrine system for treatment of hormonal  
 CC imbalances, the immune system for intervention in antigen processing,  
 CC secreted immunomodulators, and viral processing, as well as  
 CC anti-tumour applications, such as regulation of membrane trafficking  
 CC during rapid cell division.

XX Sequence 708 AA;

Query Match 42.7%; Score 41; DB 19; Length 708;  
 Best Local Similarity 38.9%; Pred. No. 80;

Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NLADRMGLSGVQEIKEQ 20

Db 373 nhqkrsigtgqldlker 390

RESULT 13

W50143

ID W50143 standard; protein; 748 AA.

XX AC W50143;

XX 28-AUG-1998 (first entry)

XX Cyanobacterial phytochrome Cphi apoprotein.

XX Phytofluor; fluorescent label; phytochrome; Cphi; cyanobacterium.

XX Synecocystis sp. strain PCC6803.

XX Key Location/Qualifiers

FT Modified-site 538

FT /note= "undetermined amino acid residue"

FT Peptide 536..544

FT /note= "histidine kinase transmitter module

FT conserved motif"

FT Peptide 642..653

FT /note= "histidine kinase transmitter module

FT conserved motif"

FT Peptide 679..687

FT /note= "histidine kinase transmitter module

FT conserved motif"

FT Peptide 694..698

FT /note= "histidine kinase transmitter module

FT conserved motif"

FT Peptide 709..714

FT /note= "histidine kinase transmitter module

FT conserved motif"

XX

PN W09805944-A1.

XX 12-FEB-1998.

XX 01-AUG-1997; 97WO-US13529.

XX 02-AUG-1996; 96US-0023217.

XX (REGC ) UNIV CALIFORNIA.

XX Lagarias JC, Murphy JT;

XX WPI: 1998-145711/13.

PT Adducts of apoprotein polypeptide and chromophore as label,  
 PT particularly for bio-molecules - used as fluorescent markers in  
 PT immunoassays, nucleic acid hybridisation, detecting protein-protein  
 PT interaction etc., are stable with high molar absorption

PS Example 4; Fig 10B; 87pp; English.

XX This polypeptide comprises cyanobacterial phytochrome 1 (Cph1) of  
 CC Synecocystis sp. PCC6803. Its amino acid sequence was deduced  
 CC from locus sir0473 genomic DNA. Expression of the 748-residue  
 CC polypeptide in *E. coli* and incubation with phycocyanobilin yields  
 CC an adduct with a red, far-red photoreversible phytochrome  
 CC signature. The invention provides a new class of fluorescent  
 CC protein adducts (designated phytofluors) that are generally suitable  
 CC for use as fluorescent markers. They comprise a protein component  
 CC (an apoprotein) and a bilin chromophore such as phycoerythrobilin.  
 CC Preferred apoproteins are obtained from plants, e.g. oats (see  
 CC W50144), from green algae, e.g. *Mesotaenium caldariorum* (see  
 CC W50145), or cyanobacteria such as *Synechocystis*. Truncated  
 CC apoproteins consisting of the N-terminal chromophore domain are  
 CC especially preferred. Recombinant apoproteins assemble  
 CC spontaneously with the bilin chromophore. Claimed compositions  
 CC comprise a protein, glycoprotein, antibody or nucleic acid to be  
 CC detected linked to the fluorescent adduct. They are used in assays  
 CC for detecting the other member of a specific binding pair, e.g.  
 CC immunoassay of antigens, immuno-histochemical labelling, as nucleic  
 CC acid probes for Southern blotting, for identification of manufactured  
 CC products, also to detect protein-protein interactions, including  
 CC studies on intracellular protein localisation and identification of  
 CC transfected cells. The phytofluors make ideal fluorescent markers  
 CC because they have a long wavelength absorption maximum and high  
 CC molar absorption coefficient, and are stable to light and pH.

XX Sequence 748 AA;

Query Match 42.7%; Score 41; DB 19; Length 748;

Best Local Similarity 57.1%; Pred. No. 86;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 NLADRMGLSGVQEQ 16

Db 354 nhpdrllgltsgq 367

RESULT 14

R03167

ID R03167 standard; protein; 332 AA.

XX AC R03167;

XX 15-AUG-1990 (first entry)

XX Amino acid sequence of uricase enzyme encoded by a gene from *Bacillus* sp.  
 DE TB-90.

XX Uricase enzyme; *Bacillus* sp. TB-90; uric acid.

XX *Bacillus*.

OS

```

XX GB2221910-A.
PN
XX
XX 21-FEB-1990.
PD
XX
XX 02-AUG-1989; 89GB-0017690.
PF
XX
XX 17-AUG-1988; 88JP-0203239.
PR
XX
XX (SAPB ) SAPPORO BREWERIES (TOYM).
PA
XX Shigyo T, Sugihara K, Takamoto Y, Takashio M, Kanimura M;
PI Yamamoto K, Kojima Y, Kikuchi T, Emi S;
XX
XX WPI; 1990-053072/08.
DR N-PSDB; Q91871.
XX
XX DNA coding for uricase enzyme contained in transformed microorganisms -
PT used as assay reagent for determ. of uric acid in blood or urine
PT
XX Disclosure; Fig 1; 26pp; English.
XX
XX New in the patent are its DNA, plasmids contg. its DNA, and transformants
CC contg. such plasmids. The transformants are useful for large-scale
CC prodn. of it. It is useful as an assay reagent for determ. of uric acid
CC in blood or urine.
XX
XX Sequence 332 AA;
SQ
Query Match 41.7%; Score 40; DB 11; Length 332;
Best Local Similarity 41.2%; Pred. No. 50;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 2 INLADRMGLSGVQEI K 18
Db 162 Initeqsglaglqlik 178
RESULT 15
B01670
ID B01670 standard; Protein; 423 AA.
XX
XX AC B01670;
XX
XX 01-SEP-2000 (first entry)
XX
XX Yeast type II topoisomerase protein fragment.
XX
XX Type II topoisomerase; identify; detect; eukaryote; intron;
KW ascomycetous yeast.
XX
XX Candida zeylanoides.
OS
XX EP997536-A2.
XX
XX 03-MAY-2000.
XX
XX 24-SEP-1999; 99EP-0118075.
XX
XX 24-SEP-1998; 98JP-0270337.
XX
XX (MARI-) MARINE BIOTECHNOLOGY INST CO LTD.
XX
XX Yamamoto S, Atsumi M, Kasai H, Hamada T, Nakamura S;
XX
XX WPI; 2000-295087/26.
DR N-PSDB; A50732.
XX
XX Identification and detection of eukaryotes by probing an
PT intron-containing region in a type II topoisomerase gene -
XX
XX Example 2; Page 15-16; 30pp; English.
PS

```

```

XX
XX The present invention relates to the identification of a eukaryote based
CC on the sequence of a type II topoisomerase gene intron containing region.
CC The method comprises amplifying the eukaryote DNA by PCR or cDNA cloning,
CC comprising a primer or probe having a nucleotide sequence encoding a
CC specific polypeptide. Those peptides used to create primers
CC for use in the method were selected by comparing the amino acid sequences
CC of the type II topoisomerase of organisms belonging to the Animalia,
CC Plantae, Protista and Fungi kingdoms. The regions containing conserved
CC amino acid sequences for the respective eukaryotes were chosen, and
CC included regions with the fewest possible degenerate codons. The method
CC is based on the identification and detection of DNA sequences in introns.
CC These non-coding regions are subject to a faster mutation rate than
CC coding regions and are therefore less conserved, making them very useful
CC in the identification of closely related organisms. The method is useful
CC in the detection and identification of eukaryotes. The present sequence
CC represents a protein encoded by a fragment of an ascomycetous yeast type
CC II topoisomerase gene, identified by the method of the invention.
XX
XX Sequence 423 AA;
SQ
Query Match 41.7%; Score 40; DB 21; Length 423;
Best Local Similarity 35.3%; Pred. No. 66;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
QY 3 NLADRMGLSLSGVQEIKE 19
Db 388 niedklllelaglnedkq 404
Search completed: April 24, 2001, 16:23:35
Job time: 153 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:35:06 ; Search time 62.1 seconds  
(without alignments)  
6.187 Million cell updates/sec

Title: US-09-171-432A-46

Perfect score: 96

Sequence: 1 KINLADRLGLSGVQEIKEQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pap.\*

2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pap.\*

3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pap.\*

4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pap.\*

5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pap.\*

6: /cgn2\_6/ptodata/2/1aa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	2227	4	US-08-475-886-2
2	96	100.0	2227	4	US-08-475-886-4
3	96	100.0	2227	4	US-08-475-886-6
4	96	100.0	2227	4	US-08-397-232-2
5	96	100.0	2227	4	US-08-397-232-4
6	42	43.8	1250	1	US-08-441-139-9
7	41	42.7	366	3	US-08-746-883-6
8	40	41.7	332	1	US-08-469-649-2
9	40	41.7	638	2	US-08-846-762-95
10	40	41.7	1207	1	US-07-951-715A-7
11	40	41.7	1207	3	US-08-459-448A-7
12	40	41.7	1207	3	US-08-459-595A-7
13	40	41.7	1207	3	US-08-459-504B-7
14	40	41.7	1207	4	US-08-459-444-7
15	40	41.7	1207	4	US-09-053-549-8
16	40	41.7	1227	1	US-08-448-170-8
17	40	41.7	1227	4	US-09-053-549-2
18	40	41.7	1227	4	US-08-961-803-9
19	40	41.7	1229	1	US-08-100-709-4
20	40	41.7	1229	1	US-08-176-855-4
21	40	41.7	1229	1	US-08-474-038-4
22	40	41.7	1229	2	US-08-779-046-4
23	40	41.7	1388	2	US-08-881-340-4
24	40	41.7	1388	2	US-08-685-576-1
25	39	40.6	528	3	US-08-904-871-5
26	39	40.6	748	3	US-08-904-871-6
27	39	40.6	748	3	US-08-904-871-13

28	38	39.6	185	2	US-08-691-814B-12	Sequence 12, Appl
29	38	39.6	261	1	US-07-940-605A-2	Sequence 2, Appl
30	38	39.6	261	1	US-08-184-422-8	Sequence 8, Appl
31	38	39.6	261	1	US-08-360-923A-2	Sequence 2, Appl
32	38	39.6	261	1	US-08-446-922-4	Sequence 4, Appl
33	38	39.6	261	2	US-08-431-055-4	Sequence 4, Appl
34	38	39.6	261	2	US-08-690-096-2	Sequence 2, Appl
35	38	39.6	261	2	US-08-249-189-12	Sequence 12, Appl
36	38	39.6	261	2	US-08-484-624A-12	Sequence 12, Appl
37	38	39.6	261	2	US-08-477-733B-12	Sequence 12, Appl
38	38	39.6	261	3	US-08-763-995-2	Sequence 2, Appl
39	38	39.6	261	3	US-09-088-913A-12	Sequence 12, Appl
40	38	39.6	261	3	US-08-589-771B-8	Sequence 8, Appl
41	38	39.6	261	5	PCT-US93-10034-4	Sequence 4, Appl
42	38	39.6	273	1	US-08-446-922-11	Sequence 11, Appl
43	38	39.6	273	2	US-08-249-189-21	Sequence 21, Appl
44	38	39.6	273	2	US-08-484-624A-21	Sequence 21, Appl
45	38	39.6	273	2	US-08-477-733B-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475.886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2  
  
Query Match 100.0%; Score 96; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 KINLADRLGLSGVQEIKEQ 20  
Db 961 KINLADRLGLSGVQEIKEQ 980  
  
RESULT 2  
US-08-475-886-4  
; Sequence 4, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475.886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18

```
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRF
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match      100.0%; Score 96; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGVQEIKEQ 20
Db 961 KINLADRLGLSGVQEIKEQ 980

RESULT 3
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRF
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match      100.0%; Score 96; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGVQEIKEQ 20
Db 961 KINLADRLGLSGVQEIKEQ 980

RESULT 4
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRF
; ORGANISM: Wild-type Human Hepatitis A Virus, Strain HM-175
US-08-397-232-2

Query Match      100.0%; Score 96; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGVQEIKEQ 20
Db 961 KINLADRLGLSGVQEIKEQ 980

RESULT 5
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRF
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

Query Match      100.0%; Score 96; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGVQEIKEQ 20
Db 961 KINLADRLGLSGVQEIKEQ 980

RESULT 6
US-08-441-139-9
; Sequence 9, Application US/08441139
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Wittrup, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```



SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441.139  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/089,997  
FILING DATE: 06-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8646  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1250 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-441-139-9

Query Match 43.8%; Score 42; DB 1; Length 1250;  
Best Local Similarity 42.1%; Pred. No. 31;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 INLADRLGLSGVQEIKEQ 20  
DB 871 LNLIDRTLGEVYVINEIKR 889

RESULT 7  
US-08-746-883-6  
Sequence 6, Application US/08746883  
Patent No. 6093693  
GENERAL INFORMATION:  
APPLICANT: Julius, Michael H., Filipp, Dominic,  
APPLICANT: Alizadeh-Khavi, Kamel  
TITLE OF INVENTION: B Cell Activation  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Blake, Cassels & Graydon  
STREET: Box 25, Commerce Court West  
CITY: Toronto  
COUNTRY: Canada  
ZIP: M5L 1A9  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage  
COMPUTER: COMPAQ, IBM PC compatible  
OPERATING SYSTEM: MS-DOS 5.1  
SOFTWARE: WORD PERFECT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/746,883  
FILING DATE: No. 6093693ember 18, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunt, John C.  
REGISTRATION NUMBER: 36,424  
REFERENCE/DOCKET NUMBER: 47841/00008  
TELEPHONE: (416) 863-4344  
TELEFAX: (416) 863-2653  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 366 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-746-883-6

Query Match 42.7%; Score 41; DB 3; Length 366;  
Best Local Similarity 63.6%; Pred. No. 11;  
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 7 RMLGLSGVQEI 17  
DB 107 RVLGISGLQEL 117

RESULT 8  
US-08-469-649-2  
Sequence 2, Application US/08469649  
Patent No. 5728562  
GENERAL INFORMATION:  
APPLICANT: Shigyo, Tatsuro  
APPLICANT: Sugihara, Kohji  
APPLICANT: Takamoto, Yoji  
APPLICANT: Takashio, Masachika  
APPLICANT: Kamimura, Minoru  
APPLICANT: Yamamoto, Kazumi  
APPLICANT: Kojima, Yoshio  
APPLICANT: Kikuchi, Toshiro  
APPLICANT: Emi, Shigenori  
TITLE OF INVENTION: AN ISOLATED URICASE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Frislauf, Holtz, Goodman, Langer &  
ADDRESSEE: Chick, P.C.  
STREET: 767 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10017-2023  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5, inch, 360 Kb Storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,649  
FILING DATE: June 6, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Barth, Richard  
REGISTRATION NUMBER: 28,180  
REFERENCE/DOCKET NUMBER: 890578CIPD  
TELEPHONE: (212) 319-4900  
TELEFAX: (212) 319-5101  
TELEX: 236268  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 332 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-649-2

Query Match 41.7%; Score 40; DB 1; Length 332;  
Best Local Similarity 41.2%; Pred. No. 14;  
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 INLADRLGLSGVQEI 18  
DB 162 LNLTEQQSGLAGLQLIK 178

RESULT 9  
US-08-846-762-95  
Sequence 95, Application US/08846762A  
Patent No. 5994072  
GENERAL INFORMATION:

APPLICANT: Lam, Joseph S.  
APPLICANT: Burrows, Lori  
APPLICANT: Charter, Deborah  
APPLICANT: de Kievit, Teresa  
TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly  
of O-Antigen in Pseudomonas Aeruginosa  
FILE REFERENCE: 6580-089  
CURRENT APPLICATION NUMBER: US/08/846.762A  
CURRENT FILING DATE: 1997-04-30  
NUMBER OF SEQ ID NOS: 100  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 95  
LENGTH: 638  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-08-846-762-95

Query Match 41.7%; Score 40; DB 2; Length 638;  
Best Local Similarity 47.4%; Pred. No. 32;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 2 INLADRMGLSGVQEIKEQ 20  
Db 515 IDLAKRMINLSGLSIKSE 533

RESULT 10  
US-07-951-715A-7  
Sequence 7, Application US/07951715A  
Patent No. 5625136  
GENERAL INFORMATION:  
APPLICANT: Koziel, Michael G.  
APPLICANT: Desai, Nalini M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttie, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/951,715A  
FILING DATE: 25-SEP-1992  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Spruill, W. Murray  
REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8615  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-951-715A-7

Query Match 41.7%; Score 40; DB 1; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 69;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 INLADRMGLSGV 14  
Db 34 INIAGRILGV 46

RESULT 11  
US-08-459-448A-7  
Sequence 7, Application US/08459448A  
Patent No. 5859336  
GENERAL INFORMATION:  
APPLICANT: Koziel, Michael G.  
APPLICANT: Desai, Nalini M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttie, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5859336artis Corporation  
STREET: Patent & Trademark Dept., 520 White Plains  
STREET: Rd., POB 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,448A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40403  
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8582  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-448A-7

Query Match 41.7%; Score 40; DB 2; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 69;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
||:|:|:|:|:|  
DB 34 INIAGRILGLGV 46

RESULT 12  
US-08-459-595A-7  
Sequence 7, Application US/08459595A  
Patent No. 6018104

GENERAL INFORMATION:  
APPLICANT: Koziel, Michael G.  
APPLICANT: Desai, Nalini M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttie, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6018104artis Corporation  
STREET: Patent & Trademark Dept., 520 White Plains  
RD., POB 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-9005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,595A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40403  
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8582  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-595A-7

Query Match 41.7%; Score 40; DB 3; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 69;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 INLADRMGLSGV 14  
||:|:|:|:|:|  
DB 34 INIAGRILGLGV 46

RESULT 13  
US-08-459-504B-7  
Sequence 7, Application US/08459504B  
Patent No. 6075185

GENERAL INFORMATION:  
APPLICANT: Koziel, Michael G.  
APPLICANT: Desai, Nalini M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramer, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.  
APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttie, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6075185artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,504B  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/459,595  
FILING DATE: 02-JUN-1995  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-504B-7

Query Match 41.7%; Score 40; DB 3; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 69;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 INLADRMGLSGV 14  
||| |||: ||  
Db 34 INIAGRILGVLGV 46

RESULT 14  
US-08-459-444-7  
; Sequence 7, Application US/08459444A  
; Patent No. 6121014  
; GENERAL INFORMATION:  
; APPLICANT: Koziel, Michael G.  
; Desai, Nalini M.  
; Lewis, Kelly S.  
; Kramer, Vance C.  
; Warren, Gregory W.  
; Evola, Stephen V.  
; Crossland, Lyle D.  
; Wright, Martha S.  
; Merlin, Ellis J.  
; Launis, Karen L.

TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED  
NUCLEIC ACID CODING SEQUENCE

NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: NO. 6121014artis Agribusiness Biotechnology Research, Inc.  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,444A  
FILING DATE: 02-Jun-1995  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-08-459-444-7

Query Match 41.7%; Score 40; DB 4; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 69;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 INLADRMGLSGV 14  
||| |||: ||  
Db 34 INIAGRILGVLGV 46

RESULT 15  
US-09-053-549-8  
; Sequence 8, Application US/09053549  
; Patent No. 6121521  
; GENERAL INFORMATION:  
; APPLICANT: Desai, Nalini  
; TITLE OF INVENTION: NO. 6121521el Insecticidal Protein and Gene  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NO. 6121521artis Corporation  
; STREET: 3054 Cornwallis Rd.  
; CITY: Research Triangle Park  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27709

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/053,549  
FILING DATE: 01-APR-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1995  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1207 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-053-549-8

Query Match 41.7%; Score 40; DB 4; Length 1207;  
Best Local Similarity 61.5%; Pred. No. 69;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 INLADRMGLSGV 14  
||| |||: ||  
Db 34 INIAGRILGVLGV 46

Search completed: April 24, 2001, 16:35:07  
Job time: 538 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:26:02 ; Search time 73.14 Seconds  
(without alignments)  
18.792 Million cell updates/sec

Title: US-09-171-432A-46

Perfect score: 96

Sequence: 1 KINLADRLGLSGVQEIKEQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_67:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	2227	1 GNNYHM	genome polyprotein
2	96	100.0	2227	1 GNNYHR	genome polyprotein
3	96	100.0	2227	1 GNNYMK	genome polyprotein
4	96	100.0	2227	1 GNNYHB	genome polyprotein
5	96	100.0	2230	1 GNNYSA	genome polyprotein
6	47	49.0	145	2 B25199	heat shock 16K pro
7	47	49.0	172	2 S33416	heat shock protein
8	44	45.8	414	2 T06303	enoyl-CoA hydratase
9	44	45.8	646	2 S72609	GTP-binding membra
10	44	45.8	653	2 G70683	probable lepA - My
11	44	45.8	788	2 S67595	hypothetical prote
12	43	44.8	145	2 B24289	heat shock protein
13	43	44.8	158	2 S64321	hypothetical prote
14	43	44.8	370	2 H70423	oxygen-independent
15	42	43.8	143	1 HKW41	heat shock protein
16	42	43.8	547	2 A40656	hypothetical prote
17	42	43.8	646	2 D81674	conserved hypothe
18	42	43.8	1035	2 S18512	cell division cont
19	42	43.8	1250	2 A39578	SSD1 protein - yea
20	41	42.7	141	2 E64368	hypothetical prote
21	41	42.7	155	2 C34965	hypothetical 17k p
22	41	42.7	155	2 S15576	ippi protein - Shi
23	41	42.7	366	1 TDMSM4	monocyte surface g
24	41	42.7	518	1 S44183	thiamin-phosphate
25	41	42.7	605	2 D83007	regulatory protein
26	41	42.7	621	2 A71516	hypothetical prote
27	41	42.7	687	2 T09051	pepa protein - Pse
28	41	42.7	708	2 JC6329	yeast secretory pr
29	41	42.7	1036	2 F82263	probable multidrug

30	41	42.7	1338	2 T40993	protein kinase cek
31	41	42.7	1652	2 T50711	complement C3 prec
32	40	41.7	143	1 HHKW48	heat shock protein
33	40	41.7	208	2 A70132	glucose inhibited
34	40	41.7	332	2 JC4535	urate oxidase (EC
35	40	41.7	448	2 C82936	signal recognition
36	40	41.7	638	2 S51266	trsg protein - Yer
37	40	41.7	739	2 S39975	stringent response
38	40	41.7	819	1 B72128	endopeptidase La (
39	40	41.7	1042	2 H70203	isoleucine--tRNA l
40	40	41.7	1151	2 T04657	hypothetical prote
41	40	41.7	1228	2 S00873	parasporal crystal
42	40	41.7	1388	2 S70633	serine/threonine-s
43	40	41.7	1388	2 S74245	serine/threonine-s
44	40	41.7	1770	2 S56221	hypothetical prote
45	39	40.6	119	2 B69554	conserved hypothet

#### ALIGNMENTS

##### RESULT 1

GNNYHM

genome polyprotein - human hepatitis A virus (strain HM-175, wild type)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core  
B; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D

C:Species: human hepatitis A virus

A:Note: host Homo sapiens (man)

C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C:Accession: A25981

R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.

J. Virol. 61, 50-59, 1987

A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with

A:Reference number: A25981; MUID:87061253

A:Accession: A25981

A:Molecule type: genomic RNA

A:Residues: 1-2227 <COH>

C:Cross-references: EMBL:M14707; NID:G329582; PIDN:AAA45465.1; PID:G329583

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltr

F:1-23/Product: coat protein 1A #status predicted <VP4>

F:24-245/Product: coat protein 1B #status predicted <VP2>

F:246-491/Product: coat protein 1C #status predicted <VP3>

F:492-791/Product: coat protein 1D #status predicted <VP1>

F:792-980/Product: core protein 2A #status predicted <C2A>

F:981-1087/Product: core protein 2B #status predicted <C2B>

F:1088-1422/Product: core protein 2C #status predicted <C2C>

F:1423-1496/Product: protein 3A #status predicted <C3A>

F:1497-1519/Product: protein 3B #status predicted <C3B>

F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>

F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 96; DB 1; Length 2227;

Best Local Similarity 100.0%; Pred. No. 2e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KINLADRLGLSGVQEIKEQ 20

Db 961 KINLADRLGLSGVQEIKEQ 980

##### RESULT 2

GNNYHR

genome polyprotein - human hepatitis A virus

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core  
NA polymerase (EC 2.7.7.48), protein 3D

C:Species: human hepatitis A virus

A:Note: host Homo sapiens (man)

C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999

C:Accession: A03903

R:Najarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van N

Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985

**RESULT**  
**GNNYHB**



A:Cross-references: EMBL:X15461

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; core protein; polyprotein

F:1-27/Product: coat protein 1A #status predicted <C1A>

F:28-249/Product: coat protein 1B #status predicted <C1B>

F:250-495/Product: coat protein 1C #status predicted <C1C>

F:496-795/Product: coat protein 1D #status predicted <C1D>

F:796-984/Product: coat protein 2A #status predicted <C2A>

F:985-1091/Product: coat protein 2B #status predicted <C2B>

F:1092-1426/Product: core protein 2C #status predicted <C2C>

F:1427-1498/Product: core protein 3A #status predicted <C3A>

F:1499-1521/Product: protein 3B #status predicted <P3B>

F:1522-1741/Product: protein 3C #status predicted <P3C>

F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 100.0%; Score 96; DB 1; Length 2230;

Best Local Similarity 100.0%; Pred. No. 2e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20

||||| ||||| ||||| ||||| |||||

Db 965 KINLADRMGLSGVQEIKEQ 984

RESULT 6

B25199

heat shock 16K protein 2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 13-Aug-1999

C:Accession: B25199

R:Jones, D.; Russnak, R.H.; Kay, R.J.; Candido, E.P.M.

J. Biol. Chem. 261, 12006-12015, 1986

A:Title: Structure, expression, and evolution of a heat shock gene locus in Caenorhabditis

A:Reference number: A92555; MUID:86304344

A:Accession: B25199

A:Molecule type: DNA

A:Residues: 1-145 <JON>

A:Cross-references: GB:M44334; NID:g156338; PIDN:AAA28071.1; PID:g156340

C:Superfamily: alpha-crystallin

Query Match

Best Local Similarity 49.0%; Score 47; DB 2; Length 145;

Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKE 18

||||| ||||| ||||| ||||| |||||

Db 65 KINLADRMGLSGVQEIKE 82

RESULT 7

S33416

heat shock protein hsp20 - nematode (Nippostrongylus brasiliensis)

C:Species: Nippostrongylus brasiliensis

C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 13-Aug-1999

C:Accession: S33416

R:Tweedie, S.; Grigg, M.E.; Ingram, L.; Selkirk, M.E.

submitted to the EMBL Data Library, April 1993

A:Description: The expression of a small heat shock homologue is developmentally regulated

A:Reference number: S33416

A:Accession: S33416

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-172 <TWE>

A:Cross-references: EMBL:X71663; NID:g297865; PIDN:CAA50655.1; PID:g297866

C:Superfamily: alpha-crystallin

Query Match

Best Local Similarity 49.0%; Score 47; DB 2; Length 172;

Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20

||||| ||||| ||||| ||||| |||||

Db 80 KVQLDDRLTVEGMQEVKTE 99

RESULT 8

T06303

enoyl-CoA hydratase homolog Fl1C18.10 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T06303

R:Bevan, M.; Terry, N.; Ardiles, W.; Buyssehaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15589

A:Accession: T06303

A:Molecule type: DNA

A:Residues: 1-414 <BEV>

A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:Fl1C18.10

A:Experimental source: cultivar Columbia; BAC clone Fl1C18

C:Genetics:

A:Gene: ATSP:Fl1C18.10

A:Map position: 4

A:Introns: 42/3; 75/2; 108/1; 134/3; 162/3; 191/1; 219/3; 260/2; 280/3; 313/3; 347/3;

Query Match

Best Local Similarity 45.8%; Score 44; DB 2; Length 414;

Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKE 19

||||| ||||| ||||| ||||| |||||

Db 260 KIELIDKYFGLDVEEIE 278

RESULT 9

S72609

GTP-binding membrane protein lepa - Mycobacterium leprae

N:Alternate names: protein B1937\_f3\_81

C:Species: Mycobacterium leprae

C>Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999

C:Accession: S72609

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Reference number: S72580

A:Accession: S72609

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-646 <SMI>

A:Cross-references: EMBL:U00016; NID:g466961; PIDN:AAA17177.1; PID:g466991

C:Genetics:

A:Gene: lepa

A:Start codon: GTG

C:Superfamily: GTP-binding membrane protein lepa; translation elongation factor Tu ho

C:Keywords: GTP binding; membrane protein; P-loop

F:45-177/Domain: translation elongation factor Tu homology <ETU>

F:51-58/Region: nucleotide-binding motif A (P-loop)

F:174-177/Region: GTP-binding NKXD motif

Query Match

Best Local Similarity 45.8%; Score 44; DB 2; Length 646;

Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGV 14

||||| ||||| ||||| ||||| |||||

Db 57 KSTLADRMQLTGV 70

RESULT 10

G70683

probable lepa - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: G70683  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: G70683  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-653 <COL>  
A;Cross-references: GB:281368; GB:AL123456; NID:g3261656; PIDN:CAB03723.1; PID:g1655655  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: lepA  
C;Superfamily: GTP-binding membrane protein lepA; translation elongation factor Tu homolog  
C;Keywords: GTP binding  
F:53-181/Domain: translation elongation factor Tu homology <ETU>  
F:59-66/Region: nucleotide-binding motif A (P-loop)  
F:178-181/Region: GTP-binding NKXD motif

Query Match 45.8%; Score 44; DB 2; Length 653;  
Best Local Similarity 71.4%; Pred. No. 25;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KINLADRLMLGLSGV 14  
I | | | | | | | | | |  
Db 65 KSTLADRLMLQLTGV 78

RESULT 11  
S67595  
hypothetical protein YDL060w - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein D2544  
C;Species: Saccharomyces cerevisiae  
C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 20-Jun-2000  
C;Accession: S67595  
R;Bloeker, H.; Brandt, P.  
A;Reference number: S67587  
A;Accession: S67595  
A;Molecule type: DNA  
A;Residues: 1-788 <BLO>  
A;Cross-references: EMBL:274108; NID:g1431062; PID:g1431063; GSPDB:GN000004; MIPS:YDL060w  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: MIPS:YDL060w  
A;Map position: 4L

Query Match 45.8%; Score 44; DB 2; Length 788;  
Best Local Similarity 61.5%; Pred. No. 31;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 MLGLSGVQEIKEQ 20  
: | | | | | | | |  
Db 159 VFGLSGVQGVDEE 171

RESULT 12  
B24289  
heat shock protein 16-1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 28-Sep-1987 #sequence\_revision 30-Sep-1993 #text\_change 29-Oct-1999  
C;Accession: B24289; S31037; T25927; T25930  
R;Russnak, R.H.; Candido, E.P.M. Mol. Cell. Biol. 5, 1268-1278, 1985  
A;Reference number: A24289; MUID:85295957  
A;Accession: B24289  
A;Molecule type: DNA

A;Residues: 1-145 <RUS>  
A;Cross-references: EMBL:K03273; NID:g156333; PIDN:AAA28068.1; PID:g156336  
A;Note: the author translated the codon GAT for residue 17 as Tyr  
R;Ray, R.J.; Russnak, R.H.; Jones, D.; Mathias, C.; Candido, E.P.M. Nucleic Acids Res. 15, 3723-3741, 1987  
A;Title: Expression of intron-containing C. elegans heat shock genes in mouse cells d y effect of heat shock on the mammalian splicing apparatus.  
A;Reference number: S31036; MUID:87231065  
A;Accession: S31037  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 41-63 <RAY>  
A;Cross-references: EMBL:M31340; NID:g156343; PIDN:AAA28073.1; PID:g552068  
R;Bradshaw, H.  
A;Description: The sequence of C. elegans cosmid T27E4.  
A;Reference number: Z20111  
A;Accession: T25927  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-145 <BRA>  
A;Cross-references: EMBL:U64837; PIDN:AA04839.1; GSPDB:GN000023; CESP:hsp-16A  
A;Experimental source: strain Bristol N2; clone T27E4  
A;Accession: T25930  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-145 <BR2>  
A;Cross-references: EMBL:U64837; PIDN:AA04842.1; GSPDB:GN000023; CESP:hsp-16A  
A;Experimental source: strain Bristol N2; clone T27E4  
C;Genetics:  
A;Gene: CESP:hsp-16A  
A;Map position: 5  
A;Introns: 42/3  
C;Superfamily: alpha-crystallin

Query Match 44.8%; Score 43; DB 2; Length 145;  
Best Local Similarity 45.0%; Pred. No. 7.3;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 KINLADRLMLGLSGVQEIKEQ 20  
| | | | | | | | | |  
Db 65 KINLADRLMLGLSGVQEIKEQ 84

RESULT 13  
S64321  
hypothetical protein YGR030c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein G4068  
C;Species: Saccharomyces cerevisiae  
C;Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 29-Oct-1999  
C;Accession: S64321  
R;Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.  
submitted to the Protein Sequence Database, May 1996  
A;Reference number: S64071  
A;Accession: S64321  
A;Molecule type: DNA  
A;Residues: 1-158 <RIE>  
A;Cross-references: EMBL:272815; NID:g1323007; PID:e243932; PID:g1323008; GSPDB:GN000  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: MIPS:YGR030c  
A;Map position: 7R

Query Match 44.8%; Score 43; DB 2; Length 158;  
Best Local Similarity 64.3%; Pred. No. 8;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KINLADRLMLGLSGV 14  
: | | | | | | | | | |  
Db 65 QINMADRLMLGLQGV 78

Db 69 KIKLDGRELKIEGQETKSE 88

Search completed: April 24, 2001, 16:26:05  
Job time: 299 sec

## RESULT 14

H70423  
oxygen-independent coproporphyrinogen III oxidase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 11-Jun-1999  
C:Accession: H70423  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: H70423  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-370 <AQF>  
A:Cross-references: GB:AE000739; MID:g2983813; PIDN:AAC07371.1; PID:g2983815; GB:AE00065  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: hemf  
C:Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 44.88; Score 43; DB 2; Length 370;  
Best Local Similarity 66.78; Pred. No. 20;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 DRMLGLSGVQEI 17

| : : | | | | : | |

Db 78 DKVFGLSGVKEI 89

## RESULT 15

HHKW41  
heat shock protein 16-41 - Caenorhabditis elegans  
N:Alternate names: heat shock protein 16 2  
C:Species: Caenorhabditis elegans  
C:Date: 25-Feb-1985 #sequence\_revision 19-Oct-1995 #text\_change 22-Jun-1999  
C:Accession: A25199; A38884; A02917  
R:Jones, D.; Russnak, R.H.; Kay, R.J.; Candido, E.P.M.  
J. Biol. Chem. 261, 12006-12015, 1986  
A:Title: Structure, expression, and evolution of a heat shock gene locus in Caenorhabditis  
A:Reference number: A92555; MUID:86304344  
A:Accession: A25199  
A:Molecule type: DNA  
A:Residues: 1-143 <JON>  
R:Candido, E.P.M.  
submitted to GenBank, November 1985  
A:Reference number: A38884  
A:Accession: A38884  
A:Molecule type: mRNA  
A:Residues: 47-143 <CAN>  
A:Cross-references: GB:X01577; MID:g6758; PIDN:CAA25732.1; PID:g780186  
R:Russnak, R.H.; Jones, D.; Candido, E.P.M.  
Nucleic Acids Res. 11, 3187-3205, 1983  
A:Title: Cloning and analysis of cDNA sequences coding for two 16 kilodalton heat shock  
A:Reference number: A93467; MUID:83220736  
A:Accession: A02917  
A:Molecule type: mRNA  
A:Residues: 'KLCFFQ', 47-143 <RUS>  
A:Cross-references: GB:K01864; MID:g156331; PIDN:AAA28065.1; PID:g156332  
A:Note: the authors translated the codon UUG for residue 46 as Phe  
A:Note: this sequence has been revised in reference A38884  
C:Superfamily: alpha-crystallin  
C:Keywords: heat shock; stress-induced protein

Query Match 43.88; Score 42; DB 1; Length 143;  
Best Local Similarity 45.08; Pred. No. 11;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20

| | | | | : | : | | |



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:33:20 ; Search time 113.97 Seconds  
(without alignments)  
20.568 Million cell updates/sec

Title: US-09-171-432a-46

Perfect score: 96

Sequence: 1 KINLADRLGLSGVQEIKEQ 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_unclassified:\*

13: sp\_vertebrate:\*

14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	1124	14	Q84780 hepatitis a
2	96	100.0	1161	14	Q05794 hepatitis a
3	96	100.0	2216	14	Q9WMA2 hepatitis a
4	96	100.0	2218	14	Q67824 hepatitis a
5	96	100.0	2218	14	Q67817 hepatitis a
6	96	100.0	2227	14	Q67825 hepatitis a
7	96	100.0	2227	14	Q67826 hepatitis a
8	96	100.0	2227	14	Q9WMA4 hepatitis a
9	96	100.0	2227	14	Q9WMA3 hepatitis a
10	96	100.0	2227	14	Q9WMA1 hepatitis a
11	96	100.0	2227	14	Q9WMA0 hepatitis a
12	96	100.0	2227	14	Q9WMA9 hepatitis a
13	96	100.0	2227	14	Q9IFH5 hepatitis a
14	71	74.0	184	14	Q87092 simian hepa
15	48	50.0	622	2	Q9RDC9 streptomyce
16	44	45.8	414	10	Q9SZ48 arabisopsis
17	44	45.8	788	3	Q07381 sacccharomyc
18	43	44.8	146	5	Q21062 caenorhabdi
19	43	44.8	211	10	Q9SFG4 arabisopsis

Q9PYV8 xestia c-nl  
Q67418 aquifex aeo  
Q93889 caenorhabdi  
Q9n691 euplotes oc  
Q9pj8 chlamydia m  
Q91vt3 arabidopsis  
Q9vid1 drosophila  
Q9rev5 carnobacter  
Q46315 carnobacter  
Q84432 chlamydia t  
Q34208 pseudomonas  
Q00471 homo sapien  
P97878 rattus norv  
Q9kt18 vibrio chol  
Q9y7n8 schizosacch  
Q90633 gallus gall  
Q27922 caenorhabdi  
Q52360 bacillus pu  
Q9qr72 kaposi's sa  
Q9lse1 arabidopsis  
Q9pr53 ureaplasma  
Q9qdl7 cyrtipedium  
P94153 alcaligenes  
P88958 kaposi's sa  
Q56919 yersinia en

20 43 44.8 232 14 Q9PYV8  
21 43 44.8 370 2 Q67418  
22 42 43.8 104 5 Q93889  
23 42 43.8 414 5 Q9n691  
24 42 43.8 646 2 Q9PJW8  
25 42 43.8 885 10 Q9LVT3  
26 41 42.7 213 5 Q9VID1  
27 41 42.7 425 2 Q9REV5  
28 41 42.7 442 2 Q46315  
29 41 42.7 621 2 Q84432  
30 41 42.7 687 2 Q34208  
31 41 42.7 708 4 Q00471  
32 41 42.7 708 11 P97878  
33 41 42.7 1036 2 Q9KT18  
34 41 42.7 1338 3 Q9Y7N8  
35 41 42.7 1652 13 Q90633  
36 40 41.7 144 5 Q27922  
37 40 41.7 145 2 Q52360  
38 40 41.7 316 4 Q9UGF7  
39 40 41.7 366 14 Q9QR72  
40 40 41.7 432 10 Q9LSE1  
41 40 41.7 448 2 Q9PRS3  
42 40 41.7 478 14 Q9QDL7  
43 40 41.7 488 2 P94153  
44 40 41.7 545 14 P88958  
45 40 41.7 638 2 Q56919

## ALIGNMENTS

RESULT 1

Q84780 ID 084780 PRELIMINARY; PRT; 1124 AA.  
AC 084780;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE RNA FOR CAPSID VP4-VP1 AND NS-PROTEINS (NON-STRUCTURAL PROTEINS)  
DE (FRAGMENT).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ovchinnikov Y.A., Sverdlov E.D., Tsarev S.A., Arsenyan S.G.,  
RA Kholina T.O., Chiznikov V.E., Petrov N.A., Prikhod'ko G.G.,  
RA Blinov V.M., Vasilenko S.K., Sandakhchiev L.S., Kusov Y.Y.,  
RA Grabko V.I., Fleer G.P., Balyan M.S., Drozdov S.G.;  
RL Dokl. Biochem. 285:379-383(1986).  
DR EMBL; X04200; CAA27797.1; -;  
DR EMBL; AL1312; CAA00953.1; -;  
KW Nonstructural protein.  
FT NON\_TER 1  
FT NON\_TER 1124 1124  
SQ SEQUENCE 1124 AA; 127026 MW; 38449E2D2ABDF8CA CRC64;

Query Match 100.0%; Score 96; DB 14; Length 1124;  
Best Local Similarity 100.0%; Pred. No. 5.3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KINLADRLGLSGVQEIKEQ 20

Db 918 KINLADRLGLSGVQEIKEQ 937

RESULT 2

Q05794 ID Q05794 PRELIMINARY; PRT; 1161 AA.  
AC Q05794; Q67800; Q67801; Q67802; Q67803; Q67804; Q67805; Q67806;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 GENOME POLYPROTEIN (COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
 P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
 P3D (EC 2.7.7.48)) (FRAGMENT).  
 Hepatitis A virus.  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 Hepatovirus.  
 NCBI\_TaxID=12092;  
 [1]  
 SEQUENCE FROM N.A.  
 Sverdlov S.D., Tsarev S.A., Markova S.V., Vasilenko S.K.,  
 Chizhikov V.E., Petrov N.A., Kusov Y.Y., Nastashenko T.A.,  
 Balayan M.S.;  
 Mol. Gen. Microbiol. Virol. 6:129-133(1987).  
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
 EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
 VP3, AND VP4.  
 EMBL; X15464; CAA33492.1; -;  
 DR INTERPRO: IPR000408; -;  
 DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_1.  
 KW Polyprotein; Coat protein; Core protein; RNA-directed RNA polymerase;  
 Hydrolase; Thiol protease.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
 FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
 FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
 FT CHAIN 795 900 CORE PROTEIN P2A.  
 FT CHAIN 901 1087 CORE PROTEIN P2B.  
 FT CHAIN 1088 >1161 CORE PROTEIN P2C.  
 FT NON\_TER 1161 1161  
 SQ SEQUENCE 1161 AA; 131131 MW; 38BE93789FEC3400 CRC64;

Query Match 100.0%; Score 96; DB 14; Length 1161;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KINLADRLGLSGVQEIKEQ 20  
 |||||  
 Db 955 KINLADRLGLSGVQEIKEQ 974

RESULT 3  
 Q9WMA2 PRELIMINARY; PRT; 2216 AA.  
 AC Q9WMA2;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE POLYPROTEIN  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AH3;  
 RA Fujiwara K.;  
 RT "Hepatitis A virus."  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB020566; BAA35104.1; -;  
 DR INTERPRO: IPR000408; -;  
 DR INTERPRO: IPR000605; -;  
 DR INTERPRO: IPR001205; -;  
 DR INTERPRO: IPR001643; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR PFAM; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICIVIRUSNS.  
 DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_1.  
 KW Polyprotein.  
 SQ SEQUENCE 2216 AA; 250209 MW; 1A9D93FEC21FBE82 CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2216;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KINLADRLGLSGVQEIKEQ 20  
 |||||  
 Db 961 KINLADRLGLSGVQEIKEQ 980

RESULT 4  
 Q67824 PRELIMINARY; PRT; 2218 AA.  
 ID Q67824;  
 AC Q67824;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE GBM/PRHK RNA.  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GBM;  
 RX MEDLINE=94076453; PubMed=8254770;  
 RA Graff J., Normann A., Feinstein S.M., Flehmig B.;  
 RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison  
 with two cell culture-adapted variants.";  
 RL J. Virol. 68:548-554(1994).  
 DR EMBL; X75214; CAA53024.1; -;  
 DR INTERPRO: IPR000408; -;  
 DR INTERPRO: IPR000605; -;  
 DR INTERPRO: IPR001205; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR PFAM; PF00910; RNA\_helicase; 1.  
 DR PROSITE; PS00626; RCCL\_2; UNKNOWN\_1.  
 FT CHAIN 1 785 P1 STRUCTURAL PROTEINS.  
 FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEINS.  
 FT CHAIN 1417 2218 P3 NONSTRUCTURAL PROTEINS.  
 SQ SEQUENCE 2218 AA; 250502 MW; CA72DF0922104C0E CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2218;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KINLADRLGLSGVQEIKEQ 20  
 |||||  
 Db 955 KINLADRLGLSGVQEIKEQ 974

RESULT 5  
 Q67817 PRELIMINARY; PRT; 2218 AA.  
 ID Q67817;  
 AC Q67817;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE COMPLETE GENOME.  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F.G.;  
 RX MEDLINE=95381623; PubMed=7653108;  
 RA Beneduce F., Pisani G., Divizia M., Pana A., Morace G.;  
 RT "Complete nucleotide sequence of a cytopathic hepatitis A virus strain  
 isolated in Italy.";  
 RL Virus Res. 36:299-309(1995).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-F.G.;  
 RA Morace G.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X83302; CAA58281.1; -;  
 DR INTERPRO; IPR000408; -;  
 DR INTERPRO; IPR000605; -;  
 DR INTERPRO; IPR001205; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR PFAM; PF00910; RNA\_helicase; 1.  
 DR PROSITE; PS00626; RCCL2; UNKNOWN\_1.  
 SQ SEQUENCE 2218 AA; 250476 MW; 813B21D3E4E533CA CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2218;  
 Best Local Similarity 100.0%; Pred. No. 1.le-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20  
 |||||  
 DB 955 KINLADRMGLSGVQEIKEQ 974

## RESULT 6

Q67825 PRELIMINARY; PRT; 2227 AA.  
 AC Q67825;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE GBM/WT RNA.  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GBM;  
 RX MEDLINE-94076453; PubMed-8254770;  
 RA Graff J., Normann A., Feinstone S.M., Flehmig B.;  
 RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison with two cell culture-adapted variants.";  
 RL J. Virol. 68:548-554(1994).  
 DR EMBL; X75215; CAA53025.1; -;  
 DR INTERPRO; IPR000408; -;  
 DR INTERPRO; IPR000605; -;  
 DR INTERPRO; IPR001205; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR PFAM; PF00910; RNA\_helicase; 1.  
 DR PROSITE; PS00626; RCCL2; UNKNOWN\_1.  
 FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.  
 FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.  
 FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.  
 SQ SEQUENCE 2227 AA; 251563 MW; 4C4D79D352F936B4 CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.le-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20  
 |||||  
 DB 961 KINLADRMGLSGVQEIKEQ 980

## RESULT 7

Q67826 PRELIMINARY; PRT; 2227 AA.  
 AC Q67826;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE GBM/HFS RNA.

OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GBM;  
 RX MEDLINE-94076453; PubMed-8254770;  
 RA Graff J., Normann A., Feinstone S.M., Flehmig B.;  
 RT "Nucleotide sequence of wild-type hepatitis A virus GBM in comparison with two cell culture-adapted variants.";  
 RL J. Virol. 68:548-554(1994).  
 DR EMBL; X75216; CAA53026.1; -;  
 DR INTERPRO; IPR000408; -;  
 DR INTERPRO; IPR000605; -;  
 DR INTERPRO; IPR001205; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR PFAM; PF00910; RNA\_helicase; 1.  
 DR PROSITE; PS00626; RCCL2; UNKNOWN\_1.  
 FT CHAIN 1 791 P1 STRUCTURAL PROTEIN.  
 FT CHAIN 792 1422 P2 NONSTRUCTURAL PROTEIN.  
 FT CHAIN 1423 2227 P3 NONSTRUCTURAL PROTEIN.  
 SQ SEQUENCE 2227 AA; 251496 MW; 48BCB7C962319457 CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.le-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20  
 |||||  
 DB 961 KINLADRMGLSGVQEIKEQ 980

## RESULT 8

Q9WMA4 PRELIMINARY; PRT; 2227 AA.  
 ID Q9WMA4;  
 AC Q9WMA4;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE POLYPROTEIN.  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AHI;  
 RA Fujiwara K.;  
 RT "Hepatitis A virus.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB020564; BAA35102.1; -;  
 DR INTERPRO; IPR000408; -;  
 DR INTERPRO; IPR000605; -;  
 DR INTERPRO; IPR001205; -;  
 DR INTERPRO; IPR001643; -;  
 DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR PFAM; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICVIRUSNS.  
 DR PROSITE; PS00626; RCCL2; UNKNOWN\_1.  
 KW Polyprotein.  
 SQ SEQUENCE 2227 AA; 251304 MW; 0DEF6D2AEC29C0CE CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.le-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGVQEIKEQ 20  
 |||||  
 DB 961 KINLADRMGLSGVQEIKEQ 980

RESULT 9  
Q9WMA3 PRELIMINARY; PRT: 2227 AA.  
ID Q9WMA3  
AC Q9WMA3;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AH2;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020565; BAA35103.1; -;  
DR INTERPRO; IPR000408; -;  
DR INTERPRO; IPR000605; -;  
DR INTERPRO; IPR001205; -;  
DR INTERPRO; IPR001643; -;  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PRINTS; PR00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSNS.  
DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2227 AA; 251440 MW; E04F846CEC7F50FD CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.le-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KINLADRLGLSGVQEIKEQ 20  
|||||  
Db 961 KINLADRLGLSGVQEIKEQ 980

RESULT 10  
Q9WMA1 PRELIMINARY; PRT: 2227 AA.  
ID Q9WMA1  
AC Q9WMA1;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PH1;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020567; BAA35105.1; -;  
DR INTERPRO; IPR000408; -;  
DR INTERPRO; IPR000605; -;  
DR INTERPRO; IPR001205; -;  
DR INTERPRO; IPR001643; -;  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSNS.  
DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2227 AA; 251415 MW; F92C8E23233FC5621 CRC64;  
Query Match 100.0%; Score 96; DB 14; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.le-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KINLADRLGLSGVQEIKEQ 20  
|||||  
Db 961 KINLADRLGLSGVQEIKEQ 980  
RESULT 11  
Q9WMA0 PRELIMINARY; PRT: 2227 AA.  
ID Q9WMA0;  
AC Q9WMA0;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FH2;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020568; BAA35106.1; -;  
DR INTERPRO; IPR000408; -;  
DR INTERPRO; IPR000605; -;  
DR INTERPRO; IPR001205; -;  
DR INTERPRO; IPR001643; -;  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSNS.  
DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 2227 AA; 251118 MW; 53E86B4432127E9B CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.le-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KINLADRLGLSGVQEIKEQ 20  
|||||  
Db 961 KINLADRLGLSGVQEIKEQ 980  
RESULT 12  
Q9WMA9 PRELIMINARY; PRT: 2227 AA.  
ID Q9WMA9;  
AC Q9WMA9;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-FH3;  
RA Fujiwara K.;  
RT "Hepatitis A virus.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB020569; BAA35107.1; -;  
DR INTERPRO; IPR000408; -;  
DR INTERPRO; IPR000605; -;  
DR INTERPRO; IPR001205; -;  
DR INTERPRO; IPR001643; -;  
DR PFAM; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PFAM; PF00910; RNA\_helicase; 1.



```

DR PRINTS: PR00918; CALICIVIRUSN.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
KW Polypeptidein.
SQ SEQUENCE 2227 AA; 251387 MW; 817640187672A23D CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1;le-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KINLADRMGLSGVQEIKEQ 20
Db 961 KINLADRMGLSGVQEIKEQ 980

RESULT 13
Q91FH5 PRELIMINARY; PRT; 2227 AA.
AC Q91FH5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE POLYPROTEIN PRECURSOR.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HAF-203;
RA Baptista M.L., Silva M., de Lima M.A., Yoshida C.F., Gaspar A.M.,
RA Pires Lopes M.Q., Galler R.;
RT "Nucleotide sequence of the HAF-203 hepatitis A virus strain isolated
RT in Brazil and expression of the VP1 gene in a bacterial system.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF268386; AAF80114.1; -.
SQ SEQUENCE 2227 AA; 251432 MW; 81913AECC9A04200 CRC64;

Query Match 100.0%; Score 96; DB 14; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1;le-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KINLADRMGLSGVQEIKEQ 20
Db 961 KINLADRMGLSGVQEIKEQ 980

RESULT 14
Q87092 PRELIMINARY; PRT; 184 AA.
AC Q87092;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE SEGMENT 2A-ENCODED PROTEIN (FRAGMENT).
OS Simian hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90012354; PubMed=2552172;
RA Brown E.A., Jansen R.W., Lemon S.M.;
RT "Characterization of a simian hepatitis A virus (HAV): antigenic and
RT genetic comparison with human HAV.";
RL J. Virol. 63:4932-4937(1989).
DR EMBL; M34085; AAA47483.1; -.
FT NON_TER 1
FT NON_TER 184
SQ SEQUENCE 184 AA; 21322 MW; 9FA851FA964A88A8 CRC64;

```

```

Query Match 74.0%; Score 71; DB 14; Length 184;
Best Local Similarity 93.3%; Pred. No. 0.00081;
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KINLADRMGLSGVQ 15
Db 170 KINLADRMGLSGVQ 184

RESULT 15
Q9RDC9 PRELIMINARY; PRT; 622 AA.
AC Q9RDC9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE GTP-BINDING PROTEIN.
GN LEPA.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL136503; CAB66240.1; -.
DR INTERPRO; IPR000795; -.
DR PFAM; PF00009; GTP_EFTU; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
SQ SEQUENCE 622 AA; 68378 MW; 83F5C76FA2A80C7C CRC64;

```

```

Query Match 50.0%; Score 48; DB 2; Length 622;
Best Local Similarity 57.9%; Pred. No. 17;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 1 KINLADRMGLSGVQEIKE 19
Db 32 KSTLADRLMLTGVVEQRQ 50

```

Search completed: April 24, 2001, 16:33:24  
Job time: 585 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:34:08 ; Search time 38.74 Seconds  
(without alignments)  
17.685 Million cell updates/sec

Title: US-09-171-432A-46

Perfect score: 96

Sequence: 1 KINLADRMGLSGVQEIKEQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	2226	1 POLG.HPAV2	P26580 hepatitis a
2	96	100.0	2226	1 POLG.HPAV4	P26581 hepatitis a
3	96	100.0	2226	1 POLG.HPAV8	P26582 hepatitis a
4	96	100.0	2227	1 POLG.HPAVH	P08617 hepatitis a
5	96	100.0	2227	1 POLG.HPAVL	P06441 hepatitis a
6	96	100.0	2227	1 POLG.HPAVM	P13901 hepatitis a
7	96	100.0	2230	1 POLG.HPAVS	P14553 simian hepa
8	47	49.0	145	1 HS12.CAEEL	P06582 caenorhabdi
9	47	49.0	172	1 HS20.NIPBR	Q07160 nippostrotr
10	44	45.8	646	1 LEPA.MYCLE	P53530 mycobacteri
11	44	45.8	653	1 LEPA.MYCTU	P71739 mycobacteri
12	43	44.8	145	1 HS11.CAEEL	P34696 caenorhabdi
13	43	44.8	158	1 YG1K.YEAST	P53218 saccharomyc
14	42	43.8	143	1 HS16.CAEEL	P06581 caenorhabdi
15	42	43.8	151	1 SDC.HALRY	P81926 halocynthia
16	42	43.8	547	1 YJDB.SALRY	P36555 salmonella
17	42	43.8	1035	1 CC68.YEAST	P32558 saccharomyc
18	42	43.8	1250	1 SSD1.YEAST	P24276 saccharomyc
19	41	42.7	141	1 Y549.METJA	Q57969 methanococc
20	41	42.7	155	1 IPPL.SHIFL	P18008 shigella fl
21	41	42.7	313	1 CBBR.RHIME	P56885 rhizobium m
22	41	42.7	366	1 CD14.MOUSE	P40386 s probable
23	41	42.7	518	1 TH14.SCHPO	P38938 schizosacch
24	41	42.7	1338	1 CEK1.SCHPO	P02513 caenorhabdi
25	40	41.7	143	1 HS17.CAEEL	P02513 caenorhabdi
26	40	41.7	208	1 GIDB.BORBU	P53363 borrelia bu
27	40	41.7	332	1 URIC.BACSB	Q45697 bacillus sp
28	40	41.7	739	1 RELA.STREQ	Q54089 streptococc
29	40	41.7	819	1 LON.CHLPN	Q929f4 chlamydia p
30	40	41.7	1042	1 SVL.BORBU	O51773 borrelia bu
31	40	41.7	1227	1 CIBB.BACTU	O85805 bacillus th
32	40	41.7	1228	1 CIBB.BACTK	P05517 bacillus th
33	40	41.7	1229	1 CIBB.BACTU	Q45739 bacillus th

34 40 41.7 1231 1 CIBD\_BACTZ  
35 40 41.7 1233 1 CIBC\_BACTM  
36 40 41.7 1770 1 RI15\_YEAST  
37 39 40.6 119 1 YO33\_ARCFU  
38 39 40.6 251 1 ASTA\_ASTFL  
39 39 40.6 365 1 AROC\_HELPJ  
40 39 40.6 365 1 AROC\_HELPJ  
41 39 40.6 454 1 SR54\_AQUAE  
42 39 40.6 482 1 HOXA\_ALCEU  
43 39 40.6 748 1 PHYL\_SYNY3  
44 39 40.6 819 1 LON\_CHLTR  
45 38.5 40.1 269 1 MIND\_GUITH

#### ALIGNMENTS

##### RESULT 1

POLG\_HPAV2  
ID POLG\_HPAV2 STANDARD; PRT: 2226 AA.  
AC P26580;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
DE P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 24a).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12094;  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination.";  
RL J. Virol. 65:2056-2065(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS.  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PPM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M59810; AAA45468.1;  
CC MEROPS: C03.005;  
CC InterPro: IPR000605;  
CC InterPro: IPR001205;  
CC Pfam: PF00680; RNA\_dep\_RNA\_pol: 1.  
CC Pfam: PF00910; RNA\_helicase: 1.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
RNA-DIRECTED POLYMERASE P3D.



Polyprotein; Coat protein; Core protein; Transferase;  
RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
KW CHAIN 1 23  
KW CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
FT VARIANT 77 77  
FT VARIANT 764 764  
FT VARIANT 821 821  
FT VARIANT 1052 1052  
FT VARIANT 1062 1062  
FT VARIANT 1118 1118  
FT VARIANT 1151 1151  
FT VARIANT 1163 1163  
FT VARIANT 1277 1277  
FT VARIANT 1500 1500  
FT VARIANT 1805 1805  
FT VARIANT 1930 1930  
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AE6740A6 CRC64;  
Query Match 100.0%; Score 96; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 KINLADRLMLGLSGVQEIKEQ 20  
Db 961 KINLADRLMLGLSGVQEIKEQ 980  
RESULT 5  
POLG\_HPAVL STANDARD; PRT; 2227 AA.  
AC P06441;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain LA).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12099;  
RN SEQUENCE FROM N.A.  
RP MEDLINE=85190549; PubMed=2986127;  
RX Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,  
RA Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus.";  
Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; K02990; AAA45472.1; -  
DR

POLG\_HPAVL STANDARD; PRT; 2227 AA.  
AC P06441; 081082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS  
P2A TO P2C; PROBABLE PROTEINS P3A TO P3C; RNA-DIRECTED RNA POLYMERASE  
P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain HM-175).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12098;  
RN SEQUENCE FROM N.A.  
RP MEDLINE=87061253; PubMed=3023706;  
RX Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,  
RA Baroudy B.M.;  
RT "Complete nucleotide sequence of wild-type hepatitis A virus:  
RT comparison with different strains of hepatitis A virus and other  
RT picornaviruses."; J. Virol. 61:50-59(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATTENUATED;  
RX MEDLINE=87175701; PubMed=3031686;  
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstein S.M.,  
RA Purcell R.H.;  
RT "Complete nucleotide sequence of an attenuated hepatitis A virus:  
RT comparison with wild-type virus."; J. Virol. 61:50-59(1987).  
RN [3]  
RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
RX MEDLINE=85166289; PubMed=2984684;  
RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,  
RA Purcell R.H., Feinstein S.M.;  
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid  
RT proteins and RNA polymerase."; J. Virol. 59:100-106(1985).  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
CC VP3, AND VP4.  
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED  
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
CC SHOWN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M14114; AAA45475.1; -  
DR EMBL; M14707; AAA45465.1; -  
DR EMBL; M14707; AAA45466.1; ALT\_INIT.  
DR EMBL; M16632; AAA45471.1; -  
DR PIR; A25981; GNNYHM.  
DR PIR; A25914; GNNYMK.  
DR PIR; A03905; A03905.  
DR MEROPS; C03.005; -  
DR InterPro; IPR000605; -  
DR InterPro; IPR001205; -  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR



```

DR EMBL; D00924; BAA00766.1; -
DR EMBL; X15461; CAA33490.1; -
DR PIR; A30470; GNYSYA.
DR PIR; S04885; S04885.
DR MEROPS; C03.005; -
DR InterPro; IPR000605; -
DR InterPro; IPR001205; -
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 1 27
FT CHAIN 28 249
FT CHAIN 250 495
FT CHAIN 496 795
FT CHAIN 796 984
FT CHAIN 985 1091
FT CHAIN 1092 1426
FT CHAIN 1427 1498
FT CHAIN 1499 1521
FT CHAIN 1522 1741
FT CHAIN 1742 2230
FT CHAIN 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;

Query Match 100.0%; Score 96; DB 1; Length 2230;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KINLADRLGLSGVQEIKEQ 20
Db 965 KINLADRLGLSGVQEIKEQ 984

RESULT 8
ID HS12_CAEEL STANDARD; PRT; 145 AA.
AC P06582;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE HEAT SHOCK PROTEIN HSP16-2.
GN HSP16-2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304344; PubMed=3017958;
RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
RT "Structure, expression, and evolution of a heat shock gene locus in
Caenorhabditis elegans that is flanked by repetitive elements.";
RL J. Biol. Chem. 261:12006-12015(1986).
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14334; AAA28071.1; -
DR PIR; B25199; B25199.
DR InterPro; IPR002068; -
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 145 AA; 16242 MW; 8A73449F99161889 CRC64;

Query Match 49.0%; Score 47; DB 1; Length 145;
Best Local Similarity 55.6%; Pred. No. 0.76;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KINLADRLGLSGVQEIKE 18
Db 65 KINLADRLGLSGVQEIKE 82.

RESULT 9
ID HS20_NIPBR STANDARD; PRT; 172 AA.
AC Q07160;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE HEAT SHOCK PROTEIN HOMOLOG (HSP20).
GN HSP20.
OS Nippostrongylus brasiliensis.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Heligmonellidae; Nippostrongylinae;
OC Nippostrongylus
OX NCBI_TaxID=27835;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94081860; PubMed=8259127;
RA Tweedie S., Grigg M.E., Ingram L., Selkirk M.E.;
RT "The expression of a small heat shock protein homologue is
developmentally regulated in Nippostrongylus brasiliensis.";
RL Mol. Biochem. Parasitol. 61:149-154(1993).
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71663; CAA50655.1; -
DR PIR; S33416; S33416.
DR InterPro; IPR002068; -
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
KW Heat shock.
SQ SEQUENCE 172 AA; 20227 MW; 2CDAA711CE60B1C0 CRC64;

Query Match 49.0%; Score 47; DB 1; Length 172;
Best Local Similarity 45.0%; Pred. No. 0.91;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KINLADRLGLSGVQEIKEQ 20
Db 80 KVLDDRLDTVEGMEVQKTE 99

RESULT 10
LEPA_MYCLE
ID LEPA_MYCLE STANDARD; PRT; 646 AA.
AC P53530;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GTP-BINDING PROTEIN LEPA.
GN LEPA OR B1937 F3.81.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;

```

```

RN SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- MISCELLANEOUS: THE OVERPRODUCTION OF THIS PROTEIN IS LETHAL TO
CC M. LEPRAE.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC LEPA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00016; AAA17177.1; -.
DR HSSP; P13551; IELO.
DR InterPro; IPR000795; -.
DR Pfam; PF00009; GTP_EFTU; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
DR GTP-binding. 51 58 GTP (BY SIMILARITY).
FT NP_BIND 120 124 GTP (BY SIMILARITY).
FT NP_BIND 174 177 GTP (BY SIMILARITY).
SQ SEQUENCE 646 AA; 71329 MW; 256DC1EAE894C4A3 CRC64;

Query Match 45.8%; Score 44; DB 1; Length 646;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGV 14
DQ 57 KSTLADRMQLQTGV 70
I | | | | | | | | | |

RESULT 11
LEPA_MYCTU
ID LEPA_MYCTU STANDARD; PRT; 653 AA.
AC P17139;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GTP-BINDING PROTEIN LEPA
GN LEPA OR RV2404C OR MCV253.16.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC LEPA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z81368; CAB03723.1; -.
DR HSSP; P13551; IELO.
DR TubercuList; RV2404C; -.
DR InterPro; IPR000795; -.
DR Pfam; PF00009; GTP_EFTU; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
DR GTP-binding. 59 66 GTP (BY SIMILARITY).
FT NP_BIND 124 128 GTP (BY SIMILARITY).
FT NP_BIND 178 181 GTP (BY SIMILARITY).
SQ SEQUENCE 653 AA; 72395 MW; DA4AFE10E6C25755 CRC64;

Query Match 45.8%; Score 44; DB 1; Length 653;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KINLADRMGLSGV 14
DQ 65 KSTLADRMQLQTGV 78
I | | | | | | | | | |

RESULT 12
HS11_CAEEL
ID HS11_CAEEL STANDARD; PRT; 145 AA.
AC P34696;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HEAT SHOCK PROTEIN HSP16-1.
GN (HSP16-1A OR T27E4.2) AND (HSP16-1B OR T27E4.8).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85295957; PubMed=4033652;
RA Russnak R.H., Candido E.P.M.;
RT "Locus encoding a family of small heat shock genes in Caenorhabditis
RT elegans: two genes duplicated to form a 3.8-kilobase inverted
RT repeat."
RL Mol. Cell. Biol. 5:1268-1278(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Bradshaw H.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 41-63 FROM N.A.
RX MEDLINE=87231065; PubMed=3588308;
RA Kay R.J., Russnak R.H., Jones D., Mathias C., Candido E.P.M.;
RT "Expression of intron-containing C. elegans heat shock genes in mouse
RT cells demonstrates divergence of 3' splice site recognition sequences
RT between nematodes and vertebrates, and an inhibitory effect of heat
RT shock on the mammalian splicing apparatus."
RL Nucleic Acids Res. 15:3723-3741(1987).
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03273; AAA28068.1; -.

```



```

DR EMBL; M31340; AAA28073.1; -
DR EMBL; U64837; AAB04842.1; -
DR EMBL; U64837; AAB04839.1; -
DR PIR; B24289; B24289.
DR WormPep; T27E4.2; CE14249.
DR WormPep; T27E4.8; CE14249.
DR InterPro; IPR002068; -
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 145 AA; 16253 MW; 06C36A1F06D15A11 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 145;
Best Local Similarity 45.0%; Pred. No. 3.5;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGVOEIKEQ 20
DB 65 KINLGHTLISIOGEELKTE 84

RESULT 13
YGIK_YEAST
ID YGIK_YEAST STANDARD; PRT; 158 AA.
AC P53218;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 18.2 KDA PROTEIN IN ERVI-GLS2 INTERGENIC REGION.
GN YGR030C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII."
RL Yeast 13:1077-1090(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72815; CAA97018.1; -
DR SGD; S0003262; POP6.
KW Hypothetical protein.
SQ SEQUENCE 159 AA; 18210 MW; 6C27A73FAD521181 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 158;
Best Local Similarity 64.3%; Pred. No. 3.8;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGV 14
DB 65 QINMADRSIGLQOV 78

RESULT 14
HS16_CAEEL
ID HS16_CAEEL STANDARD; PRT; 143 AA.
AC P06581; P02514;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)

```

```

DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HEAT SHOCK PROTEIN HSP16-41.
GN HSP16-41.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304344; PubMed=3017958;
RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
RT "Structure, expression, and evolution of a heat shock gene locus in
Caenorhabditis elegans that is flanked by repetitive elements."
RL J. Biol. Chem. 261:12006-12015(1986).
RN [2]
RP SEQUENCE OF 47-143 FROM N.A.
RX MEDLINE=83220736; PubMed=6190129;
RA Russnak R.H., Jones D., Candido E.P.M.;
RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
heat shock proteins (hsp) in Caenorhabditis elegans: homology with
the small hsp of Drosophila."
RL Nucleic Acids Res. 11:3187-3205(1983).
CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14334; AAA28070.1; ALT_SEQ.
DR EMBL; X01577; CAA25732.1; -
DR PIR; A38884; HHKW41.
DR PIR; A25199; A25199.
DR InterPro; IPR002068; -
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 143 AA; 16252 MW; CID0F59D26E36C24 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 143;
Best Local Similarity 45.0%; Pred. No. 5.1;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 KINLADRLGLSGVOEIKEQ 20
DB 69 KIKLDGRELKIEGIOETKSE 88

RESULT 15
SODC_HALRO
ID SODC_HALRO STANDARD; PRT; 151 AA.
AC P81926;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SUPEROXIDE DISMUTASE [CU-ZN] (EC 1.15.1.1).
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyridae; Halocynthia.
OX NCBI_TaxID=7729;
RN [1]
RP SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Hemocyte, and Plasma;
RX MEDLINE=99302489; PubMed=10374259;
RA Abe Y., Ishikawa G., Satoh H., Azumi K., Yokosawa H.;
RT "Primary structure and function of superoxide dismutase from the
ascidian Halocynthia roretzi."
RL Comp. Biochem. Physiol. 122B:321-326(1999).

```

```

CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS. THE PLASMA SUPEROXIDE
CC DISMUTASE HAS PHAGOCYTOSIS-STIMULATING ACTIVITY AND MAY PLAY AN
CC IMPORTANT ROLE IN THE BIOLOGICAL DEFENSES OF THE ORGANISM.
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).
CC -!- ENZYME REGULATION: INHIBITED BY KCN AND DIETHYLTHIOCARBAMATE.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- PTM: ACYLATED N-TERMINAL SERINE.
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR InterPro: IPR001424;
DR PRINTS: PR00068; CUZNDISMUTASE.
DR PROSITE: PS00087; SOD_CU_ZN_1; 1.
DR PROSITE: PS00332; SOD_CU_ZN_2; FALSE_NEG.
KW Oxidoreductase; Copper; Zinc.
FT METAL 43 43 COPPER (BY SIMILARITY).
FT METAL 45 45 COPPER (BY SIMILARITY).
FT METAL 60 60 COPPER AND ZINC (BY SIMILARITY).
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 77 77 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 117 117 COPPER (BY SIMILARITY).
FT DISULFID 54 143 BY SIMILARITY.
SQ SEQUENCE 151 AA; 15489 MW; 162F181A82275AF0 CRC64;

Query Match 43.8% Score 42; DB 1; Length 151;
Best Local Similarity 43.8%; Pred. No. 5.4;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 INLADRLGLSGVQEI 17
Db 94 VNITDRMISLTGEHSI 109

```

Search completed: April 24, 2001, 16:34:10  
Job time: 566 sec